

# Similarity and bisimilarity notions appropriate for characterizing indistinguishability in fragments of the calculus of relations

George H.L. Fletcher\*    Marc Gyssens†    Dirk Leinders†  
 Jan Van den Bussche‡    Dirk Van Gucht‡  
 Stijn Vansummeren§

## Abstract

Motivated by applications in databases, this paper considers various fragments of the calculus of binary relations. The fragments are obtained by leaving out, or keeping in, some of the standard operators, along with some derived operators such as set difference, projection, coprojection, and residuation. For each considered fragment, a characterization is obtained for when two given binary relational structures are indistinguishable by expressions in that fragment. The characterizations are based on appropriately adapted notions of simulation and bisimulation.

## 1 Introduction

The calculus of relations [26, 12, 19, 25] consists of five natural operations on binary relations: union, intersection, complementation, composition, and converse. These operators can be applied to given binary relations, combined with the four standard constant relations: empty, full, identity, and diversity. The calculus of relations is a very natural formalism and occurs within logics for reasoning about binary relations, notably dynamic and description logics [15, 5]. The calculus also has motivated the development of the theory of relation algebras [20, 17]. In the present paper, however, we are not looking at abstract relation algebras, but rather at the question of *indistinguishability* of two given finite binary relational structures within the calculus of relations.

Indistinguishability of structures in various logics is one of the most basic concepts studied in finite model theory and in the study of the expressive power

---

\*Eindhoven University of Technology, The Netherlands (g.h.l.fletcher@tue.nl)

†Hasselt University and Transnational University of Limburg, Belgium ({marc.gyssens, dirk.leinders, jan.vandenbussche}@uhasselt.be)

‡Indiana University, Bloomington, USA (vgucht@cs.indiana.edu)

§Université libre de Bruxelles, Belgium (stijn.vansummeren@ulb.ac.be)

of database query languages [8, 18, 3]. Indeed the calculus of relations, as a core relational algebra query language on binary relations, is very relevant to the field of databases. Binary relations, or, equivalently, directed graphs, show up naturally in data on the Web [2, 10], dataspace [11], Linked Data [7, 16], and RDF data [1]. Moreover, in restriction to directed graphs that are trees, the relational calculus is closely tied to the XML query language XPath, and the expressive power of XPath and various fragments has been intensively investigated [6, 22, 23, 14].

Here, working with general finite binary relation structures rather than trees, we consider, in addition to the five binary relation operations and four constant binary relations mentioned above, also four derived operations that are well known in the literature: set difference; projection; coprojection; and residuation. These derived operations can be expressed in terms of the other operations and constants, but can still be interesting on their own when considering fragments where some other operations or constants have been left out. We consider set difference because it is the standard domain-independent alternative to complementation in database query languages [3]. We consider projection and coprojection (existential and universal quantification) because they are standard logical operations, and have been shown important in the XPath setting, so it is natural to study their behaviour when generalising from trees to general graphs. Finally, we consider residuation because it is similar to the standard relational division operation in databases, and corresponds to the set containment join [21]. Obviously, one could keep on inventing additional operations on binary relations and study their interdependencies, but we hope our chosen set of operations is not too large and well-motivated.

Our goal now is to understand the relative importance of the various operations and the effect of their presence on indistinguishability. Thereto we consider all possible fragments of the calculus of relations that can be constructed as follows. The most basic fragment we consider has the empty and identity relations as constants, and the operations union, composition, and intersection. Then all other fragments arise by adding any choice of the remaining operations and constants. For each fragment, we provide a characterization of when two finite binary relation structures are indistinguishable by expressions in the fragment. Our approach is to come up with bisimilarity-like characterizations [28].

To conclude this Introduction, we note another motivation to understand indistinguishability in database query language fragments, apart from the intrinsic foundational motivation. This is the new approach of *structural* indexing to database query processing, proposed by some of us and others [9, 29], whereby a given query expression is processed by accessing blocks of data indistinguishable by the operations used in the given expression. By the results of our work, these blocks can be computed using similarity or bisimilarity checks.

This paper is organized as follows. In Section 2 we define the language fragments formally, and define the notion of indistinguishability. In Section 3 we discuss different ways how indistinguishability can be characterized; in particular we discuss the connection with multi-dimensional modal logics, and the 3-variable fragment of first-order logic. In Section 4, we consider the frag-

ments with the set difference operation. In Section 5, we consider the fragments without set difference. In Section 6 we show that indistinguishability of finite structures is decidable in polynomial time. We conclude in Section 7.

## 2 Language fragments and indistinguishability

We assume an infinite universe of atomic data elements, denoted by  $U$ . A *binary relation* on  $U$  is a subset of  $U^2 = U \times U$ . We further fix an arbitrary finite set  $\Lambda$  of *relation names*, called the *vocabulary*. In the calculus of relations, a *structure* is then a pair  $\mathcal{G} = (V, (R^\mathcal{G})_{R \in \Lambda})$  where  $V$  is a subset of  $U$  and each  $R^\mathcal{G}$  is a binary relation on  $V$ . The set  $V$  is called the set of nodes of  $\mathcal{G}$ ; the vocabulary  $\Lambda$  can be thought of as a set of edge labels whereby  $\mathcal{G}$  can be thought of as an edge-labeled directed graph. When  $V$  is finite, the structure is said to be a *finite structure*.

Expressions in the calculus of relations are built recursively from the relation names  $R$ , and the constant symbols empty (0), all (1), diversity (0'), and identity (1'), using the following standard and/or derived operations. The standard operations are union ( $e_1 \cup e_2$ ), intersection ( $e_1 \cap e_2$ ), complementation ( $e^c$ ), composition ( $e_1 \circ e_2$ ), and converse ( $e^{-1}$ ); the derived operations we consider are set difference ( $e_1 - e_2$ ), projection ( $\pi_1 e$  or  $\pi_2 e$ ), co-projection ( $\bar{\pi}_1 e$  or  $\bar{\pi}_2 e$ ), left residual ( $e_1 / e_2$ ) and right residual ( $e_1 \setminus e_2$ ).<sup>1</sup>

Semantically, on any structure  $\mathcal{G}$  as above, an expression  $e$  defines a binary relation, denoted by  $e(\mathcal{G})$ . For convenience, we recall the semantics of the constants and the standard operations.

$$\begin{aligned}
R(\mathcal{G}) &= R^\mathcal{G}; \\
0(\mathcal{G}) &= \emptyset; \\
1(\mathcal{G}) &= V^2; \\
0'(\mathcal{G}) &= \{(s, t) \mid s, t \in V \ \& \ s \neq t\}; \\
1'(\mathcal{G}) &= \{(s, s) \mid s \in V\}; \\
(e_1 \cup e_2)(\mathcal{G}) &= e_1(\mathcal{G}) \cup e_2(\mathcal{G}); \\
(e_1 \cap e_2)(\mathcal{G}) &= e_1(\mathcal{G}) \cap e_2(\mathcal{G}); \\
e^c(\mathcal{G}) &= \{(s, t) \mid s, t \in V \ \& \ (s, t) \notin e(\mathcal{G})\}; \\
(e_1 \circ e_2)(\mathcal{G}) &= \{(s, t) \mid (\exists v)((s, v) \in e_1(\mathcal{G}) \ \& \ (v, t) \in e_2(\mathcal{G}))\}; \\
(e^{-1})(\mathcal{G}) &= \{(s, t) \mid (t, s) \in e(\mathcal{G})\}.
\end{aligned}$$

---

<sup>1</sup>To distinguish between set difference and the right residual, we use the minus sign ( $-$ ) for set difference.

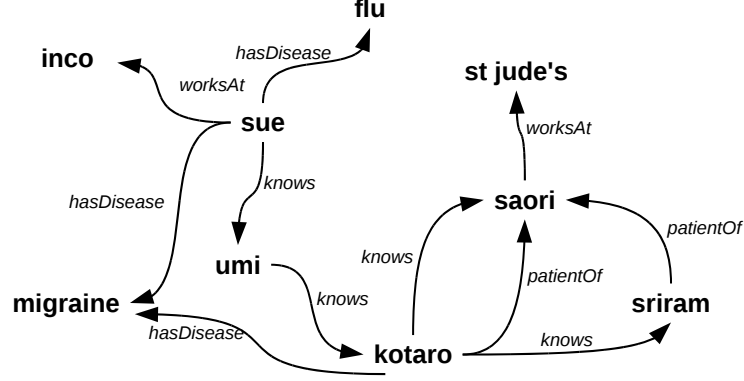


Figure 1: Example structure from Example 1.

The semantics of the derived operations is as follows:

$$\begin{aligned}
(e_1 - e_2)(\mathcal{G}) &= \{(s, t) \mid (s, t) \in e_1(\mathcal{G}) \ \& \ (s, t) \notin e_2(\mathcal{G})\} \\
\pi_1(e)(\mathcal{G}) &= \{(s, s) \mid (\exists t)(s, t) \in e(\mathcal{G})\} \\
\pi_2(e)(\mathcal{G}) &= \{(s, s) \mid (\exists t)(t, s) \in e(\mathcal{G})\} \\
\bar{\pi}_1(e)(\mathcal{G}) &= \{(s, s) \mid s \in V \ \& \ \neg(\exists t)(s, t) \in e(\mathcal{G})\} \\
\bar{\pi}_2(e)(\mathcal{G}) &= \{(s, s) \mid s \in V \ \& \ \neg(\exists t)(t, s) \in e(\mathcal{G})\} \\
(e_1 / e_2)(\mathcal{G}) &= \{(s, t) \mid (\forall v)((t, v) \in e_2(\mathcal{G}) \rightarrow (s, v) \in e_1(\mathcal{G}))\} \\
(e_1 \setminus e_2)(\mathcal{G}) &= \{(s, t) \mid (\forall v)((v, s) \in e_1(\mathcal{G}) \rightarrow (v, t) \in e_2(\mathcal{G}))\}
\end{aligned}$$

**Example 1.** Figure 1 shows a finite structure  $\mathcal{G}$ . The set of nodes equals  $\{\text{migraine}, \text{flu}, \text{sue}, \text{umi}, \text{saori}, \text{sriram}, \text{st jude's}, \text{inco}\}$ , and the vocabulary  $\Lambda$  equals  $\{\text{knows}, \text{worksAt}, \text{patientOf}, \text{hasDisease}\}$ .

- The doctors (i.e., persons having patients), can be retrieved from  $\mathcal{G}$  by the expression

$$e_1 = \pi_2(\text{patientOf})$$

resulting in  $e_1(\mathcal{G}) = \{(\text{saori}, \text{saori})\}$ .

- The people and the doctors they know can be obtained by the expression

$$e_2 = \text{knows} \circ e_1$$

resulting in  $e_2(\mathcal{G}) = \{(\text{kotaro}, \text{saori})\}$ .

- The doctors and the hospitals where they practice:

$$(e_1 \circ \text{worksAt})(\mathcal{G}) = \{(\text{saori}, \text{st jude's})\}.$$

- Ill people without medical care:

$$(\pi_1(\text{hasDisease}) - \pi_1(\text{patientOf}))(\mathcal{G}) = \{(\text{sue}, \text{sue})\}.$$

- Healthy doctors:

$$(e_1 \cap \bar{\pi}_1(\text{hasDisease}))(\mathcal{G}) = \{(\text{saori}, \text{saori})\}.$$

- Finally, the doctors who know all the patients of some other doctor can be retrieved by the expression

$$e_1 \cap \pi_1((\text{knows} / \text{patientOf}) \cap 0'),$$

which on our example graph yields the empty relation, since the graph contains only one doctor.

**Equivalence** Two expressions  $e_1$  and  $e_2$  are *equivalent*, denoted by  $e_1 \equiv e_2$ , if  $e_1(\mathcal{G}) = e_2(\mathcal{G})$  for all possible structures  $\mathcal{G}$ . The following equivalences demonstrate that the derived operations are indeed derived, and also present some additional interdependencies among the constants and operations considered in this paper:

$$\begin{aligned} 1 &\equiv 0^c \\ 0' &\equiv 1'^c \\ e_1 - e_2 &\equiv e_1 \cap e_2^c \\ e^c &\equiv 1 - e \\ \pi_1(e) &\equiv (e \circ e^{-1}) \cap 1' \equiv (e \circ (0' \cup 1')) \cap 1' \equiv \bar{\pi}_1(\bar{\pi}_1(e)) \\ \pi_2(e) &\equiv (e^{-1} \circ e) \cap 1' \equiv ((0' \cup 1') \circ e) \cap 1' \equiv \bar{\pi}_2(\bar{\pi}_2(e)) \\ \bar{\pi}_i(e) &\equiv 1' - \pi_i(e) \\ e_1 / e_2 &\equiv (e_1^c \circ e_2^{-1})^c \equiv (e_1^{-1} \setminus e_2^{-1})^{-1} \\ e_1 \setminus e_2 &\equiv (e_1^{-1} \circ e_2^c)^c \equiv (e_2^{-1} / e_1^{-1})^{-1} \end{aligned}$$

**Language fragments** We will consider various fragments of the calculus of relations. The most basic fragment we consider is denoted by  $\mathcal{C}$ : it has the constants 0 and 1' and the operators composition, union, and intersection. All other fragments are defined by adding to  $\mathcal{C}$  some additional constants and operators. The fragment  $\mathcal{C}(\pi)$ , for example, consists of the expressions built up from the relation names, 0, and 1', using the operations composition, union, intersection, and projection.<sup>2</sup> The full calculus of relations corresponds to the fragment

<sup>2</sup>We only consider fragments containing both the first and second projection ( $\pi_1$  and  $\pi_2$ ) or none of them. Similarly for coprojection, and also similarly, we only consider fragments containing both the left and right residual ( $/$  and  $\setminus$ ) or none of them.

$\mathcal{C}^{(-1, c)}$ , since all other operations can be derived in it. We will characterize indistinguishability for all fragments that contain  $\mathcal{C}$ .

**Indistinguishability** A *marked structure*  $\bar{\mathcal{G}}$  is a pair  $(\mathcal{G}, a, b)$  where  $\mathcal{G}$  is a relational structure, and  $(a, b)$  is an ordered pair of nodes from  $\mathcal{G}$ . Let  $\mathcal{F}$  be a fragment of the calculus of relations. The  $\mathcal{F}$ -*type* of  $\bar{\mathcal{G}}$ , denoted by  $\text{tp}_{\mathcal{F}}(\bar{\mathcal{G}})$ , is defined as the set of all expressions  $e \in \mathcal{F}$  such that  $(a, b) \in e(\mathcal{G})$ . For two marked structures  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\bar{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$ , we write  $\bar{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}} \bar{\mathcal{G}}_2$  if  $\text{tp}_{\mathcal{F}}(\bar{\mathcal{G}}_1) \subseteq \text{tp}_{\mathcal{F}}(\bar{\mathcal{G}}_2)$ , i.e., for every expression  $e \in \mathcal{F}$  such that  $(a_1, b_1) \in e(\mathcal{G}_1)$ , also  $(a_2, b_2) \in e(\mathcal{G}_2)$ . We then say that  $\bar{\mathcal{G}}_2$  is *one-sided indistinguishable* from  $\bar{\mathcal{G}}_1$  in  $\mathcal{F}$ . When both  $\bar{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}} \bar{\mathcal{G}}_2$  and  $\bar{\mathcal{G}}_2 \Rightarrow_{\mathcal{F}} \bar{\mathcal{G}}_1$ , we say that  $\bar{\mathcal{G}}_1$  and  $\bar{\mathcal{G}}_2$  are *indistinguishable in  $\mathcal{F}$*  and denote this by  $\bar{\mathcal{G}}_1 \equiv_{\mathcal{F}} \bar{\mathcal{G}}_2$ .

Since indistinguishability is the same as one-sided indistinguishability in both directions, it is more general to look for a characterization of one-sided indistinguishability, and indeed we will provide such characterizations whenever we can. On the other hand, we will verify in Proposition 4 that for any fragment where set difference is present, indistinguishability actually coincides with one-sided indistinguishability (except in a trivial case). For these fragments, we will thus just talk about indistinguishability for short.

In our characterizations of indistinguishability we will also refer to the *atomic*  $\mathcal{F}$ -type of  $\bar{\mathcal{G}}$ . The *atomic* expressions are those from the finite set  $\text{Atom} = \{1', 0', 1, 0\} \cup \{R, R^{-1} \mid R \in \Lambda\}$ . The atomic  $\mathcal{F}$ -type of  $\bar{\mathcal{G}}$ , denoted by  $\text{atp}_{\mathcal{F}}(\bar{\mathcal{G}})$ , then equals  $\text{Atom} \cap \text{tp}_{\mathcal{F}}(\bar{\mathcal{G}})$ . The set of atomic expressions belonging to the fragment  $\mathcal{F}$  will be denoted by  $\text{ats}(\mathcal{F})$ . Note that  $\text{atp}_{\mathcal{F}}(\bar{\mathcal{G}})$  is always a subset of  $\text{ats}(\mathcal{F})$ .

**Degrees and paths** It is customary to parameterize characterizations of indistinguishability by the degree of expressions. For an expression  $e$ , we define the *degree*  $\deg(e)$  of  $e$  as follows. Every atomic expression has degree zero. Then,

$$\begin{aligned} \deg(e_1 \cup e_2) &= \deg(e_1 \cap e_2) = \deg(e_1 - e_2) = \max(\deg(e_1), \deg(e_2)); \\ \deg(e^c) &= \deg(e^{-1}) = \deg(e); \\ \deg(e_1 \circ e_2) &= \deg(e_1 / e_2) = \deg(e_1 \setminus e_2) = 1 + \max(\deg(e_1), \deg(e_2)); \\ \deg(\pi_1(e)) &= \deg(\pi_2(e)) = \deg(\bar{\pi}_1(e)) = \deg(\bar{\pi}_2(e)) = 1 + \deg(e). \end{aligned}$$

The degree of an expression is the maximum depth of nested applications of the composition, projection, co-projection, and the left and right residual operation. Intuitively, the degree corresponds to the quantifier rank of the obvious translation of  $e$  into first-order logic.

For a fragment  $\mathcal{F}$  of the calculus of relations and a natural number  $k$ , we denote the set of expressions in  $\mathcal{F}$  of degree at most  $k$  by  $\mathcal{F}_k$ .

**Definition 2** ( $\mathcal{F}$ - $k$ -path). Let  $\mathcal{F}$  be a fragment of the calculus of relations, and let  $k$  be a natural number.

We define the expressions  $\text{paths}_k^{\mathcal{F}}$  by induction on  $k$  as follows:

$$\begin{aligned}\text{paths}_0^{\mathcal{F}} &:= \bigcup_{e \in \text{ats}(\mathcal{F})} e \\ \text{paths}_{k+1}^{\mathcal{F}} &:= \text{paths}_k^{\mathcal{F}} \cup (\text{paths}_k^{\mathcal{F}} \circ \text{paths}_k^{\mathcal{F}}).\end{aligned}$$

Since  $\mathcal{F}$  is by definition closed under union and composition,  $\text{paths}_k^{\mathcal{F}}$  is in  $\mathcal{F}_k$ . By definition,  $\text{paths}_k^{\mathcal{F}}$  is equivalent to the union of all compositions of at most  $2^k$  atomic expressions in  $\mathcal{F}$ . It is instructive to note that, for the most basic fragment  $\mathcal{C}$ , given a structure  $\mathcal{G}$ , a pair  $(a, b)$  is in  $\text{paths}_k^{\mathcal{C}}(\mathcal{G})$  if and only if there is a directed path of length at most  $2^k$  between  $a$  and  $b$  in  $\mathcal{G}$  viewed as a graph. Likewise for the fragment  $\mathcal{C}^{(-1)}$ , but then for undirected paths. Note also that when  $1$ , or just  $0'$ , is in  $\mathcal{F}$ , then  $1 = 1' \cup 0'$  is always a subexpression of  $\text{paths}_k^{\mathcal{F}}$ , so that  $\text{paths}_k^{\mathcal{F}}$  becomes equivalent to  $1$ . Thus, when  $1$  or  $0'$  is in  $\mathcal{F}$  then for any structure  $\mathcal{G}$  and any  $k$  we have  $\text{paths}_k^{\mathcal{F}} = V^2$  where  $V$  is the node set of  $\mathcal{G}$ .

The following lemma shows the relevance of  $\text{paths}_k^{\mathcal{F}}$ .

**Lemma 3.** *Let  $\mathcal{F}$  be a fragment of the calculus of relations and let  $\mathcal{G}$  be a structure. For any expression  $e \in \mathcal{F}_0$ , we have  $e(\mathcal{G}) \subseteq \text{paths}_0^{\mathcal{F}}(\mathcal{G})$ . Furthermore, unless the residual operations  $/$  and  $\backslash$  are present in  $\mathcal{F}$ , for any natural number  $k$  and any expression  $e \in \mathcal{F}_k$ , we have  $e(\mathcal{G}) \subseteq \text{paths}_k^{\mathcal{F}}(\mathcal{G})$ .*

*Proof.* By structural induction on  $e$ . If  $e$  is atomic, then  $e$  has degree zero and clearly  $e(\mathcal{G}) \subseteq \text{paths}_0^{\mathcal{F}}(\mathcal{G})$  as  $\text{paths}_0^{\mathcal{F}}(\mathcal{G})$  is the union of all atomic expressions in  $\mathcal{F}$ .

If  $e$  is  $e_1 \cup e_2$ ,  $e_1 \cap e_2$ , or  $e_1 - e_2$ , the result follows immediately from the induction hypothesis.

If  $e$  is  $\pi_1(e_1)$ ,  $\pi_2(e_1)$ ,  $\bar{\pi}_1(e_1)$ , or  $\bar{\pi}_2(e_1)$ , the result is immediate because  $\pi_1(e_1)(\mathcal{G}) \subseteq 1'(\mathcal{G}) \subseteq \text{paths}_k^{\mathcal{F}}(\mathcal{G})$ . (Similarly for  $\pi_2(e_1)$ ,  $\bar{\pi}_1(e_1)$ , and  $\bar{\pi}_2(e_1)$ .)

If  $e$  is  $e_1^c$ , the result is trivial. Indeed, if complementation is present in  $\mathcal{F}$ , then  $0' = 1'^c$  as well, and since  $1' \cup 0' = 1$ , we have  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}) = 1(\mathcal{G})$ . Hence,  $\text{paths}_k^{\mathcal{F}}(\mathcal{G}) = 1(\mathcal{G})$  since  $\text{paths}_k^{\mathcal{F}}(\mathcal{G}) \supseteq \text{paths}_0^{\mathcal{F}}(\mathcal{G})$ . The result now trivially follows as  $e(\mathcal{G}) \subseteq 1(\mathcal{G})$  for any expression  $e$ .

If  $e$  is  $e_1^{-1}$ , first observe that, if the converse operation is present in  $\mathcal{F}$ , then  $(b, a) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G})$  implies  $(a, b) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G})$  as is readily verified by induction on  $k$ . Now, assume  $(a, b) \in e_1^{-1}(\mathcal{G})$ . Then,  $(b, a) \in e_1(\mathcal{G})$ , and, by induction,  $(b, a) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G})$  whence  $(a, b) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G})$ .

Finally, if  $e$  is  $e_1 \circ e_2$ , let  $k_1 = \deg(e_1)$ ,  $k_2 = \deg(e_2)$ , and  $\ell = \max(k_1, k_2)$ . Note that  $k = \ell + 1$ . Now assume  $(a, b) \in e_1 \circ e_2(\mathcal{G})$ . Then, for some  $c \in V$ , we have  $(a, c) \in e_1(\mathcal{G})$  and  $(c, b) \in e_2(\mathcal{G})$ . By induction, we have  $(a, c) \in \text{paths}_{k_1}^{\mathcal{F}}(\mathcal{G})$  and  $(c, b) \in \text{paths}_{k_2}^{\mathcal{F}}(\mathcal{G})$ . Since  $\text{paths}_i^{\mathcal{F}}(\mathcal{G}) \subseteq \text{paths}_{i+1}^{\mathcal{F}}(\mathcal{G})$  for any  $i$ , we have  $(a, c) \in \text{paths}_{\ell}^{\mathcal{F}}(\mathcal{G})$  and  $(c, b) \in \text{paths}_{\ell}^{\mathcal{F}}(\mathcal{G})$ . Hence  $(a, b) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G})$  as desired.  $\square$

Note the exception made in the above lemma for the residual operations: indeed, for example, the degree-one expression  $0/0$  is equivalent to  $1$ , whereas

in general we do not have  $1(\mathcal{G}) \subseteq \text{paths}_1^{\mathcal{C}}(\mathcal{G})$ . (Note that  $\text{paths}_k^{\mathcal{C}(\cdot, \setminus)}$  is the same as  $\text{paths}_k^{\mathcal{C}}$ .)

To conclude this section, we show, as promised earlier:

**Proposition 4.** *Let  $\mathcal{F}$  be a fragment of the calculus of relations where set difference or complementation is present. Let  $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be two marked structures, and let  $k$  be a natural number. Consider the equivalence*

$$\overline{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}_k} \overline{\mathcal{G}}_2 \quad \Leftrightarrow \quad \overline{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \overline{\mathcal{G}}_2.$$

*This equivalence holds in each of the following cases:*

1. *either complementation,  $0'$ , or  $1$  is present in  $\mathcal{F}$ ;*
2.  *$k > 0$  and the residuals are present in  $\mathcal{F}$ ;*
3.  *$(a_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$ .*

*If none of the above cases is satisfied, then  $\overline{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}_k} \overline{\mathcal{G}}_2$  hold trivially and  $\overline{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \overline{\mathcal{G}}_2$  holds if and only if  $(a_2, b_2) \notin \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$ .*

*Proof.* Let us first argue the claim for when none of the three cases holds. Given the absence of cases 2 and 3, we obtain from Lemma 3 that  $(a_1, b_1)$  cannot belong to  $e(\mathcal{G}_1)$  for any  $e \in \mathcal{F}_k$ . Hence,  $\overline{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}_k} \overline{\mathcal{G}}_2$  is voidlessly satisfied. Moreover, clearly  $\overline{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \overline{\mathcal{G}}_2$  iff  $(a_2, b_2)$  does not belong to  $e(\mathcal{G}_1)$ , for any  $e \in \mathcal{F}_k$ , either. We now note that the latter holds iff  $(a_2, b_2) \notin \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$ . Indeed, the only-if is clear since  $\text{paths}_k^{\mathcal{F}}$  belongs to  $\mathcal{F}_k$ ; the if-direction is again given by Lemma 3.

In order to prove the equivalence claim, assume  $\overline{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}_k} \overline{\mathcal{G}}_2$ ; we must show that  $\overline{\mathcal{G}}_2 \Rightarrow_{\mathcal{F}_k} \overline{\mathcal{G}}_1$ . Thereto, let  $e \in \mathcal{F}_k$  such that  $(a_2, b_2) \in e(\mathcal{G}_2)$ ; we must show that  $(a_1, b_1) \in e(\mathcal{G}_1)$ . We consider the three different cases from the statement of the proposition.

1. If either complementation,  $0'$ , or  $1$  is present, then the complement  $e^c$  of an expression  $e$  of degree  $k$  is expressible by an expression of degree  $k$ . Indeed, if complementation is present, this is trivial; if  $0'$  is present then we have  $e^c \equiv (1' \cup 0') - e$ ; and if  $1$  is present then we have  $e^c \equiv 1 - e$ . (Note that  $\mathcal{F}$  has complement or set difference, so if complement is not present, the set difference is.)  
Now assume, for the sake of contradiction, that  $(a_1, b_1) \notin e(\mathcal{G}_1)$ . Then  $(a_1, b_1) \in e^c(\mathcal{G}_1)$ , whence  $(a_2, b_2) \in e^c(\mathcal{G}_2)$  (by  $\overline{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}_k} \overline{\mathcal{G}}_2$ ), whence  $(a_2, b_2) \notin e(\mathcal{G}_2)$  which yields the desired contradiction.
2. We have already noted that  $0/0 \equiv 1$ . Note that the degree of  $0/0$  equals one. Hence, if  $k > 0$  and the residuals are present, then  $e^c$  can be expressed as  $0/0 - e$  which is still of degree  $k$ . We can now reason as in the previous case.
3. Recall that  $\text{paths}_k$  is an expression of degree  $k$ . Now suppose  $(a_{1,1}) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$ , and assume again for the sake of contradiction that  $(a_1, b_1) \notin$



$e(\mathcal{G}_1)$ . Thus  $(a_1, b_1) \in (\text{paths}_k^{\mathcal{F}} - e)(\mathcal{G}_1)$ , whence  $(a_2, b_2) \in (\text{paths}_k^{\mathcal{F}} - e)(\mathcal{G}_2)$  by  $\mathcal{G}_1 \Rightarrow_{\mathcal{F}_k} \mathcal{G}_2$ . We obtain again the desired contradiction to the effect that  $(a_2, b_2) \notin e(\mathcal{G}_2)$ .

□

### 3 Approaches to bisimilarity

Before characterizing indistinguishability for fragments of the calculus of relations, let us first look at characterizations of the full calculus. Tarski and Givant showed that the calculus has equal expressive power as  $\text{FO}^3$ , the 3-variable fragment of first-order logic [27]. For  $\text{FO}^3$ , we have the 3-pebble Ehrenfeucht-Fraïssé game as a characterization [8, 18]. Marx and Venema, however, showed that the 3-variable fragment of first-order logic has also equal expressive power as arrow logic [24], a branch of multi-dimensional modal logic devised to provide a formalization for simple reasoning about objects that are thought of as arrows. By this correspondence, bisimulations in terms of back-and-forth conditions that are well known from modal logic can be used to characterize fragments of  $\text{FO}^3$ , and, hence, of the calculus of relations.

Concretely, the language of arrow logic is a modal language with the dyadic operator  $\circ$ , the monadic operator  $\otimes$ , and the modal constant  $\text{id}$ . Formulas in arrow logic are built up from a set of propositional variables and the modal constant  $\text{id}$ , using the operators  $\circ$  and  $\otimes$ , and the boolean connectives  $\wedge$ ,  $\vee$ ,  $\neg$ . Using propositional variables to denote edge labels; by interpreting the modal constant  $\text{id}$  as being true for pairs  $(a, a)$  of identical nodes; by interpreting the monadic operator  $\otimes$  as being true for pairs  $((b, a), (a, b))$  of “arrows” such that the first arrow is the converse of the second arrow; and finally, by interpreting the dyadic operator  $\circ$  as being true for triples  $((a, b), (a, c), (c, b))$  of arrows such that the first one is obtained by composing the second and the third arrow, we can apply the characterization theorem of modal logic to immediately obtain a characterization for the full calculus of relations. We will next make this more precise.

The notion of bisimulation for multi-dimensional modal logic, specialized to the above interpretation of arrow logic, becomes the following:

**Definition 5.** Let  $\mathcal{G}_1$  and  $\mathcal{G}_2$  be two structures with node sets  $V_1$  and  $V_2$ , respectively. A non-empty relation  $Z \subseteq V_1^2 \times V_2^2$  is a *bisimulation* between  $\mathcal{G}_1$  and  $\mathcal{G}_2$  if it satisfies the following conditions:<sup>3</sup>

**Atoms** if  $(a_1, b_1, a_2, b_2)$  is in  $Z$ , then  $(a_1, b_1) \in R(\mathcal{G}_1)$  if and only if  $(a_2, b_2) \in R(\mathcal{G}_2)$ , for all  $R \in \Lambda$ ;

---

<sup>3</sup>The attentive reader will notice that the converse-forth condition and the converse-back condition are identical. This is a consequence of the symmetry of the converse operator. We could have simplified the definition by removing one of the identical conditions, but preferred to stay in line with the general format of bisimulation conditions for multidimensional modal logic.

**Forth** if  $(a_1, b_1, a_2, b_2) \in Z$ , then

- composition**( $\circ$ ) for each  $c_1 \in V_1$  there exist  $c_2 \in V_2$  such that both  $(a_1, c_1, a_2, c_2)$  and  $(c_1, b_1, c_2, b_2)$  are in  $Z$ ;
- identity**(id) if  $a_1 = b_1$  then  $a_2 = b_2$ ;
- converse**( $\otimes$ )  $(b_1, a_1, b_2, a_2) \in Z$ ;

**Back** if  $(a_1, b_1, a_2, b_2)$  is in  $Z$ , then

- composition**( $\circ$ ) for each  $c_2 \in V_2$  there exist  $c_1 \in V_1$  such that both  $(a_1, c_1, a_2, c_2)$  and  $(c_1, b_1, c_2, b_2)$  are in  $Z$ ;
- identity**(id) if  $a_2 = b_2$  then  $a_1 = b_1$ ;
- converse**( $\otimes$ )  $(b_1, a_1, b_2, a_2) \in Z$ ;

A marked structure  $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  is said to be *bisimilar* to a marked structure  $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  if there is a bisimulation  $Z$  between  $\mathcal{G}_1$  and  $\mathcal{G}_2$  containing  $(a_1, b_1, a_2, b_2)$ .

The following characterization can now be proved in an analogous way to known results in modal logic [13]:

**Proposition 6.** *Consider the full calculus of relations  $\mathcal{F} = \mathcal{C}(\circ, -^1)$ . Let  $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be finite marked structures. Then*

$$\overline{\mathcal{G}}_1 \equiv_{\mathcal{F}} \overline{\mathcal{G}}_2 \iff \overline{\mathcal{G}}_1 \text{ is bisimilar to } \overline{\mathcal{G}}_2.$$

In database theory [3], it is common to replace the complementation operator by the “safe” difference operator and, to compensate for this weaker operation, add the diversity operator. So, it is interesting to consider the fragment  $\mathcal{F}_{\text{safe}} = \mathcal{C}(-, -^1, 0')$ , with which we deal later. Furthermore we can consider the “positive fragments”, without the difference operator, which yields the fragment  $\mathcal{F}_{\text{safe}}^+ = \mathcal{C}(-^1, 0')$ .

For  $\mathcal{F}_{\text{safe}}^+$ , the above characterization can be easily adapted. It suffices in the definition of bisimulation to remove the Back condition (thus obtaining a kind of *simulation* rather than bisimulation), and add the following part to the forth-condition:

**diversity**(di) if  $a_1 \neq b_1$ , then  $a_2 \neq b_2$ .

We can then analogously show that  $(\mathcal{G}_1, a_1, b_1) \Rightarrow_{\mathcal{F}_{\text{safe}}^+} (\mathcal{G}_2, a_2, b_2)$  if and only if there exists a simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  containing  $(a_1, b_1, a_2, b_2)$ .

But if we now add to  $\mathcal{F}_{\text{safe}}^+$  the coprojection operators, however, it is no longer so easy to obtain a characterization of indistinguishability. Indeed, the coprojection operator, as a non-monotonic operator, cannot be expressed as a modality in the sense of modal logic. Another difficulty arises when we remove, e.g., the converse operator or the diversity relation. As expressions in the calculus always return pairs  $(a, b)$  of nodes such that there is a path from  $a$  to  $b$  in the graph formed by the atomic steps, it does not suffice to remove the

converse-forth or the diversity-forth parts in the definition of bisimulation; we also need to adapt the composition-forth part.

In the rest of this section, as a concrete illustration of how to deal with these difficulties, we will define a notion of similarity that characterizes indistinguishability in  $\mathcal{C}(\bar{\pi})$ , i.e.,  $\mathcal{F}_{\text{safe}}^+$  with coprojection added and diversity and converse removed. This illustration will serve as representative example for the later sections where we characterize indistinguishability for all fragments. For fragments that contain the difference operation, we will define an appropriate general notion of bisimilarity in Section 4. For fragments without difference, we will define a general notion of similarity in Section 5.

### 3.1 Indistinguishability in $\mathcal{C}(\bar{\pi})$

We begin by defining the appropriate similarity notion. Note that we define similarity *up to a certain depth*  $k$ . That is because an expression  $e$  in  $\mathcal{C}(\bar{\pi})$  of a fixed degree  $k$ , can only output pairs of nodes between which there is a path of length at most  $2^k$  (cf. Lemma 3).<sup>4</sup>

**Definition 7** ( $\mathcal{C}(\bar{\pi})$ -similarity). Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\bar{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be marked structures with node sets  $V_1$  and  $V_2$ , respectively. We say that  $\bar{\mathcal{G}}_1$  is  $\mathcal{C}(\bar{\pi})$ -similar to  $\bar{\mathcal{G}}_2$  up to depth  $k$ , denoted  $\bar{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} \bar{\mathcal{G}}_2$ , if the following conditions are satisfied:

**Atoms** if  $a_1 = b_1$ , then  $a_2 = b_2$ ; furthermore, if  $(a_1, b_1) \in R(\mathcal{G}_1)$ , then  $(a_2, b_2) \in R(\mathcal{G}_2)$ , for all  $R \in \Lambda$ ;

**Composition Forth** Only required when  $k > 0$ . For every  $c_1 \in V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_{k-1}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_1)$ , there exists  $c_2 \in V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_{k-1}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2)$  such that both  $(\mathcal{G}_1, a_1, c_1) \preceq_{k-1}^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \preceq_{k-1}^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, c_2, b_2)$ ;

**Coprojection Forth** Only required when  $k > 0$  and  $a_1 = b_1$  (whence  $a_2 = b_2$  by the Atoms condition). For every  $c_2 \in V_2$  with  $(a_2, c_2)$  in  $\text{paths}_{k-1}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(a_1, c_1)$  in  $\text{paths}_{k-1}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_1)$  such that  $(\mathcal{G}_2, a_2, c_2) \preceq_{k-1}^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_1, a_1, c_1)$ . Furthermore, for every  $c_2 \in V_2$  with  $(c_2, a_2)$  in  $\text{paths}_{k-1}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(c_1, a_1)$  in  $\text{paths}_{k-1}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_1)$  such that  $(\mathcal{G}_2, c_2, a_2) \preceq_{k-1}^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_1, c_1, a_1)$ .

Note how, in the Coprojection Forth condition, the direction of the similarity is reversed.

It is instructive to already note the following property, although it will also follow from our later results:

**Proposition 8.** *If  $\bar{\mathcal{G}}_1 \preceq_{k+1}^{\mathcal{C}(\bar{\pi})} \bar{\mathcal{G}}_2$  then  $\bar{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} \bar{\mathcal{G}}_2$ .*

<sup>4</sup>The sets  $\text{paths}_k^{\mathcal{C}(\bar{\pi})}$  and  $\text{paths}_k^{\mathcal{C}}$  are equal for any  $k$ , and consist of the ordered pairs connected by a directed path of length at most  $2^k$ .

*Proof.* There are two cases. If  $(a_1, b_1) \in \text{paths}_k(\mathcal{G}_1)$ , then in the Composition Forth condition for  $\bar{\mathcal{G}}_1 \preceq_{k+1}^{\mathcal{C}(\bar{\pi})} \bar{\mathcal{G}}_2$  we can take  $c_1 = a_1$  and obtain  $c_2 \in V_2$  such that  $(\mathcal{G}_1, a_1, a_1) \preceq_k^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, a_1, b_1) \preceq_k^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, c_2, b_2)$ . By the Atoms condition for  $(\mathcal{G}_1, a_1, a_1) \preceq_k^{\mathcal{C}(\bar{\pi})} \mathcal{G}_2(a_2, c_2)$ , we obtain  $c_2 = a_2$  and thus  $(\mathcal{G}_1, a_1, b_1) \preceq_k^{\mathcal{C}(\bar{\pi})} \mathcal{G}_2(a_2, b_2)$  as desired.

If  $(a_1, b_1) \notin \text{paths}_k(\mathcal{G}_1)$ , then  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1) \preceq_k^{\mathcal{C}(\bar{\pi})} \bar{\mathcal{G}}_2$  is voidly satisfied.  $\square$

We now show that similarity is sufficient for one-sided indistinguishability up to degree  $k$ : (recall that for any fragment  $\mathcal{F}$ , the set of expressions in  $\mathcal{F}$  of degree at most  $k$  is denoted by  $\mathcal{F}_k$ )

**Lemma 9** (Invariance lemma for  $\mathcal{C}(\bar{\pi})$ ). *If  $\bar{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} \bar{\mathcal{G}}_2$  then  $\bar{\mathcal{G}}_1 \Rightarrow_{\mathcal{C}(\bar{\pi})_k} \bar{\mathcal{G}}_2$ .*

*Proof.* Let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\bar{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  and let  $V_1$  and  $V_2$  be the node sets of the structures  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , respectively. Assume  $\bar{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} \bar{\mathcal{G}}_2$ . We prove by structural induction that, for each expression  $e$  in  $\mathcal{C}(\bar{\pi})_k$ , if  $(a_1, b_1) \in e(\mathcal{G}_1)$ , then  $(a_2, b_2) \in e(\mathcal{G}_2)$ .

We only show the reasoning for the case where  $e$  is a coprojection. So, assume  $(a_1, b_1) \in \bar{\pi}_1(e_1)(\mathcal{G}_1)$ . Then,  $a_1 = b_1$  and there does not exist  $c_1$  in  $V_1$  such that  $(a_1, c_1) \in e_1(\mathcal{G}_1)$ . By the atoms condition, we have that  $a_2 = b_2$ . To show that  $(a_2, a_2) \in \bar{\pi}_1(e_1)(\mathcal{G}_2)$ , it remains to be shown that there does not exist  $c_2$  in  $V_2$  such that  $(a_2, c_2) \in e_1(\mathcal{G}_2)$ . Suppose, for the purpose of contradiction, that such  $c_2$  does exist. Note that  $e_1$  has degree at most  $k-1$ , so, by Lemma 3, we have  $(a_2, c_2) \in \text{paths}_{k-1}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2)$ . Hence, by the Coprojection Forth condition, we have  $(\mathcal{G}_2, a_2, c_2) \preceq_{k-1}^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_1, a_1, c_1)$ . By the induction hypothesis,  $(a_1, c_1) \in e_1(\mathcal{G}_1)$ , but this contradicts  $(a_1, b_1) \in \bar{\pi}_1(e_1)(\mathcal{G}_1)$ .

The case where  $e$  is of the form  $\bar{\pi}_2(e_1)$  is analogous.  $\square$

In order to show that similarity is also necessary for indistinguishability, the following lemma is crucial.

**Lemma 10** (Representation lemma for  $\mathcal{C}(\bar{\pi})$ ). *Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  be a marked structure with  $(a_1, b_1) \in \text{paths}_k^{\mathcal{C}(\bar{\pi})}$ . Then there exists an expression  $e_{\bar{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), k}$  in  $\mathcal{C}(\bar{\pi})_k$  such that, for every structure  $\mathcal{G}_2$ :*

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), k}(\mathcal{G}_2) = \{(a_2, b_2) \in \text{paths}_k^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2) \mid \bar{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)\}.$$

*Proof.* The construction of the required expression is by induction on  $k$ . For the base of the construction we put<sup>5</sup>

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), 0} := \text{paths}_0^{\mathcal{C}(\bar{\pi})} \cap \chi_{\bar{\mathcal{G}}_1, \text{atoms}}^{\mathcal{C}(\bar{\pi})}$$

---

<sup>5</sup>In this and in many later proofs, empty intersections should be interpreted as vanishing from the larger expression; empty unions should be interpreted as the expression  $\emptyset$ .

where

$$\chi_{\overline{\mathcal{G}}_1, \text{atoms}}^{\mathcal{C}(\bar{\pi})} := \bigcap_{e \in \text{atp}_{\mathcal{C}(\bar{\pi})}(\overline{\mathcal{G}}_1)} e.$$

In the inductive step of the construction we define

$$e_{\overline{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), k+1} := \text{paths}_{k+1}^{\mathcal{C}(\bar{\pi})} \cap \chi_{\overline{\mathcal{G}}_1, \text{atoms}}^{\mathcal{C}(\bar{\pi})} \cap \varphi_{\overline{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{C}(\bar{\pi}), k+1} \cap \chi_{\overline{\mathcal{G}}_1, \text{coprojection forth}}^{\mathcal{C}(\bar{\pi}), k+1}$$

where (with  $V_1$  the set of nodes of  $\mathcal{G}_1$ )

$$\begin{aligned} \chi_{\overline{\mathcal{G}}_1, \text{atoms}}^{\mathcal{C}(\bar{\pi})} &:= \bigcap_{e \in \text{atp}_{\mathcal{C}(\bar{\pi})}(\overline{\mathcal{G}}_1)} e; \\ \varphi_{\overline{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{C}(\bar{\pi}), k+1} &:= \bigcap_{\substack{c_1 \in V_1 \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_1) \\ (c_1, b_1) \in \text{paths}_k^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_1)}} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{C}(\bar{\pi}), k} \circ e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{C}(\bar{\pi}), k}; \end{aligned}$$

and the Coprojection Forth expression is equal to 1 if  $a_1 \neq b_1$  (the Coprojection Forth condition is then indeed vacuously satisfied); if  $a_1 = b_1$  it is defined as

$$\chi_{\overline{\mathcal{G}}_1, \text{coprojection forth}}^{\mathcal{C}(\bar{\pi}), k+1} := \bar{\pi}_1 \left( \bigcup_{\substack{e \in \mathcal{C}(\bar{\pi})_k \\ (a_1, b_1) \in \bar{\pi}_1(e)(\mathcal{G}_1)}} e \right) \cap \bar{\pi}_2 \left( \bigcup_{\substack{e \in \mathcal{C}(\bar{\pi})_k \\ (a_1, b_1) \in \bar{\pi}_2(e)(\mathcal{G}_1)}} e \right).$$

Although the expressions given for the two Forth conditions are in principle infinite, they are equivalent to finite expressions. Indeed, note that the Composition Forth expression is an intersection of expressions of degree at most  $k+1$ , and that the Coprojection Forth expression involves a union over expressions of degree at most  $k$ . For any fixed  $k$ , there are only a finite number of expressions of degree  $k$  up to equivalence. As such, the infinite intersections and unions can be equivalently expressed as finite intersections and unions.

To show the correctness of the above expression, first note that  $e_{\overline{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), k}$  indeed has degree at most  $k$ . It remains to be shown that, for any structure  $\mathcal{G}_2$ ,

$$e_{\overline{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), k}(\mathcal{G}_2) = \{(a_2, b_2) \in \text{paths}_k^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2) \mid \overline{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)\}.$$

This can be shown by induction on  $k$ . For  $k = 0$  it is clear that  $(a_2, b_2) \in \chi_{\overline{\mathcal{G}}_1, \text{atoms}}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2)$  iff the Atoms condition for  $\overline{\mathcal{G}}_1 \preceq_0^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)$  is satisfied. The equivalence between  $(a_2, b_2) \in \varphi_{\overline{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{C}(\bar{\pi}), k+1}(\mathcal{G}_2)$  and the Composition Forth condition for  $\overline{\mathcal{G}}_1 \preceq_{k+1}^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)$  is also clear, assuming the induction hypothesis for  $k$ . So, in our reasoning below, we can focus on the expression corresponding to the Coprojection Forth condition.

To prove the  $\supseteq$ -direction of the asserted equality, let  $(a_2, b_2) \in \text{paths}_{k+1}^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2)$  with  $\overline{\mathcal{G}}_1 \preceq_{k+1}^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)$ . We must show that  $(a_2, b_2) \in \chi_{\overline{\mathcal{G}}_1, \text{coprojection forth}}^{\mathcal{C}(\bar{\pi}), k+1}(\mathcal{G}_2)$ . If  $a_1 \neq b_1$  then this is trivial, so we may assume  $a_1 = b_1$ , in which case also

$a_2 = b_2$  by the Atoms condition. Let  $V_2$  denote the set of nodes of  $\mathcal{G}_2$ . The expression  $\chi_{\overline{\mathcal{G}}_1, \text{coprojection forth}}^{\mathcal{C}(\bar{\pi}), k+1}$  is an intersection of two coprojections; we focus on the first coprojection, as the second one is dealt with in a similar way. Now suppose, for the sake of contradiction, that there exists  $c_2 \in V_2$  and  $e \in \mathcal{C}(\bar{\pi})_k$  with  $(a_1, b_1) \in \bar{\pi}_1(e)(\mathcal{G}_1)$  and  $(a_2, c_2) \in e(\mathcal{G}_2)$ . Since  $\overline{\mathcal{G}}_1 \preceq_{k+1}^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)$ , we also have  $(a_2, b_2) \in \bar{\pi}_1(e)(\mathcal{G}_2)$ . But then no  $c_2$  can exist so that  $(a_2, c_2) \in e(\mathcal{G}_2)$  and we have a contradiction.

For the  $\subseteq$ -direction, let  $(a_2, b_2)$  be in  $e_{\overline{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), k+1}(\mathcal{G}_2)$ . It is clear that  $(a_2, b_2) \in \text{paths}_{k+1}^{\mathcal{C}(\bar{\pi})}$  and that the Atoms condition and Composition Forth conditions are satisfied. To argue for the Coprojection Forth condition, assume  $a_1 = b_1$ , and let  $c_2 \in V_2$  such that  $(a_2, c_2) \in \text{paths}_k^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_2)$ . Suppose, for the sake of contradiction, that there does not exist  $c_1 \in V_1$  with  $(a_1, c_1) \in \text{paths}_k^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_1)$  such that  $(\mathcal{G}_2, a_2, c_2) \preceq_k^{\mathcal{C}(\bar{\pi})} (\mathcal{G}_1, a_1, c_1)$ , or, by induction, such that  $(a_1, c_1) \in e_{(\mathcal{G}_2, a_2, c_2)}^{\mathcal{C}(\bar{\pi}), k}(\mathcal{G}_1)$ . Thus  $(a_1, a_1) = (a_1, b_1) \in \bar{\pi}_1(e_{(\mathcal{G}_2, a_2, c_2)}^{\mathcal{C}(\bar{\pi}), k})(\mathcal{G}_1)$ . Since that expression is of degree at most  $k$ , and since  $(a_2, b_2) \in \chi_{\overline{\mathcal{G}}_1, \text{coprojection forth}}^{\mathcal{C}(\bar{\pi}), k+1}(\mathcal{G}_2)$ , we also have  $(a_2, a_2) = (a_2, b_2) \in \bar{\pi}_1(e_{(\mathcal{G}_2, a_2, c_2)}^{\mathcal{C}(\bar{\pi}), k})(\mathcal{G}_2)$ . Hence, there does not exist  $c$  with the property that  $(a_2, c) \in e_{(\mathcal{G}_2, a_2, c_2)}^{\mathcal{C}(\bar{\pi}), k}(\mathcal{G}_2)$ , which is a contradiction, since  $c = c_2$  does have that property. The reasoning for the  $\bar{\pi}_2$  part of the Coprojection Forth condition is entirely analogous.  $\square$

We obtain the desired result:

**Theorem 11.**  $\overline{\mathcal{G}}_1 \Rightarrow_{\mathcal{C}(\bar{\pi})_k} \overline{\mathcal{G}}_2$  if and only if  $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} \overline{\mathcal{G}}_2$ .

*Proof.* We have already seen the if-direction in Invariance Lemma 9. For the only-if direction, we use the Representation Lemma 10. Let  $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$ . Assume for the moment that  $(a_1, b_1) \in \text{paths}_k^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_1)$ . Then, since trivially  $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} \overline{\mathcal{G}}_1$ , we have  $(a_1, b_1) \in e_{\overline{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), k}(\mathcal{G}_1)$ , whence  $(a_2, b_2) \in e_{\overline{\mathcal{G}}_1}^{\mathcal{C}(\bar{\pi}), k}(\mathcal{G}_2)$ , whence  $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} \overline{\mathcal{G}}_2$  as desired. If  $(a_1, b_1) \notin \text{paths}_k^{\mathcal{C}(\bar{\pi})}(\mathcal{G}_1)$ , then  $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{C}(\bar{\pi})} \overline{\mathcal{G}}_2$  is voidly satisfied.  $\square$

## 4 Bisimilarity for fragments with set difference

### 4.1 Fragments without the residuals

In this section,  $\mathcal{F}$  is an arbitrary fixed fragment of the calculus of relations in which set difference or complementation is present, but the residual operations are not.

**Definition 12** (Bisimilarity excl. residuals). Let  $k$  be a natural number and let  $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be marked structures with node sets  $V_1$  and  $V_2$ , respectively. We say that  $\overline{\mathcal{G}}_1$  is  $\mathcal{F}$ -bisimilar to  $\overline{\mathcal{G}}_2$  up to depth  $k$ , denoted  $\overline{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ , if the following conditions are satisfied:

**Atoms**  $\text{atp}_{\mathcal{F}}(\overline{\mathcal{G}}_1) = \text{atp}_{\mathcal{F}}(\overline{\mathcal{G}}_2)$ ;

**Composition Forth** if  $k > 0$ , then, for every  $c_1$  in  $V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ , there exists  $c_2$  in  $V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$  such that both  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

**Projection Forth** if  $\pi$  is in  $(\mathcal{F})$ , if  $k > 0$ , and if  $a_1 = b_1$ , then, for every  $c_1$  in  $V_1$  with  $(a_1, c_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$  (resp.,  $(c_1, a_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ ), there exists  $c_2$  in  $V_2$  with  $(a_2, c_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$  (resp.,  $(c_2, a_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$ ) such that  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  (resp.,  $(\mathcal{G}_1, c_1, a_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ );

**Composition Back** if  $k > 0$ , then, for every  $c_2$  in  $V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$ , there exists  $c_1$  in  $V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$  such that both  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

**Projection Back** if  $\pi$  is in  $(\mathcal{F})$ , if  $k > 0$ , and if  $a_1 = b_1$ , then, for every  $c_2$  in  $V_2$  with  $(a_2, c_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$  (resp.,  $(c_2, a_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$ ), there exists  $c_1$  in  $V_1$  with  $(a_1, c_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$  (resp.,  $(c_1, a_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ ) such that  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  (resp.,  $(\mathcal{G}_1, c_1, a_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ ).

The reader may wonder why there are no Coprojection Forth and Back conditions in the above definition. The reason is that, since difference is present in the fragment, coprojection is present if and only if projection is, by the equivalences  $\bar{\pi}_i(e) \equiv 1' - \pi_i(e)$  and  $\pi_i(e) \equiv \bar{\pi}_i(\bar{\pi}_i(e))$ . Thus, it suffices to have Projection Forth and Back conditions.

**Lemma 13** (Invariance lemma—bisimilarity excl. residuals). *If  $\overline{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$  then  $\overline{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \overline{\mathcal{G}}_2$ .*

*Proof.* Let  $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  and let  $V_1$  and  $V_2$  be the node sets of the structures  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , respectively. Let  $e$  be an expression in  $\mathcal{F}_k$ . We prove by induction on the structure of  $e$  that  $(a_1, b_1) \in e(\mathcal{G}_1)$  if and only if  $(a_2, b_2) \in e(\mathcal{G}_2)$ .

For the base case where  $e$  is an atomic expression, the result follows immediately from the atoms condition in the definition of bisimilarity.

If  $e$  is  $e_1 \cup e_2$ ,  $e_1 \cap e_2$ ,  $e_1 - e_2$ , or  $e_1^c$ , the result follows immediately from the induction hypothesis.

For the case where  $e$  is  $e_1 \circ e_2$ , consider the only-if, i.e., assume that  $(a_1, b_1) \in e(\mathcal{G}_1)$ . By definition of composition, there exists  $c_1$  in  $V_1$  with  $(a_1, c_1) \in e_1(\mathcal{G}_1)$  and  $(c_1, b_1) \in e_2(\mathcal{G}_1)$ . Since  $e_1$  and  $e_2$  have depth at most  $k-1$ , by Lemma 3, both  $(a_1, c_1)$  and  $(c_1, b_1)$  are in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ . By the Composition Forth condition, there exists  $c_2$  in  $V_2$  such that both  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ . By induction, we have  $(a_2, c_2) \in e_1(\mathcal{G}_2)$  and  $(c_2, b_2) \in e_2(\mathcal{G}_2)$ , whence  $(a_2, b_2) \in e_1 \circ e_2(\mathcal{G}_2)$ . The argument for the if-direction is similar, using Composition Back instead of Composition Forth.

Now let  $e$  be of the form  $e_1^{-1}$ . By a straightforward argument using induction on  $k$ , one can verify for fragments  $\mathcal{F}$  in which the converse operator is present,

that  $(\mathcal{G}_1, a_1, b_1) \simeq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)$  implies  $(\mathcal{G}_1, b_1, a_1) \simeq_k^{\mathcal{F}} (\mathcal{G}_2, b_2, a_2)$ . We thus obtain by induction that  $(a_1, b_1) \in e(\mathcal{G}_1)$  iff  $(b_1, a_1) \in e_1(\mathcal{G}_1)$  iff  $(b_2, a_2) \in e_1(\mathcal{G}_2)$  iff  $(a_2, b_2) \in e(\mathcal{G}_2)$  as desired.

For the case where  $e$  is  $\pi_1(e_1)$ , consider the only-if, i.e., assume  $(a_1, b_1) \in \pi_1(e_1)(\mathcal{G}_1)$ . By definition of projection, we have  $a_1 = b_1$ , and there exists  $c_1$  in  $V_1$  with  $(a_1, c_1) \in e_1(\mathcal{G}_1)$ . By Lemma 3, we have  $(a_1, c_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ . By the Projection Forth condition, there exists  $c_2$  in  $V_2$  such that  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$ . By induction, we have  $(a_2, c_2) \in e_1(\mathcal{G}_2)$ , whence  $(a_2, b_2) \in \pi_1(e_1)(\mathcal{G}_2)$ . The argument for the if-direction is similar, using Projection Back instead of Composition Forth. The argument for the case where  $e$  is  $\pi_2(e_1)$  is analogous.  $\square$

**Lemma 14** (Representation lemma—bisimilarity excl. residuals). *Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  be a marked structure with  $(a_1, b_1) \in \text{paths}_k^{\mathcal{F}}$ . Then there exists an expression  $e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}$  in  $\mathcal{F}_k$  such that for every structure  $\mathcal{G}_2$ :*

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_2) = \{(a_2, b_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \mid \bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)\}.$$

*Proof.* The construction of the required expression is by induction on  $k$ . For the base of the construction we put

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{F},0} := (\text{paths}_0^{\mathcal{F}} \cap \varphi_{\bar{\mathcal{G}}_1, \text{posatoms}}^{\mathcal{F}}) - \varphi_{\bar{\mathcal{G}}_1, \text{negatoms}}^{\mathcal{F}}$$

where

$$\varphi_{\bar{\mathcal{G}}_1, \text{posatoms}}^{\mathcal{F}} := \bigcap_{e \in \text{atp}_{\mathcal{F}}(\bar{\mathcal{G}}_1)} e \quad \text{and} \quad \varphi_{\bar{\mathcal{G}}_1, \text{negatoms}}^{\mathcal{F}} := \bigcup_{e \in \text{ats}(\mathcal{F}) - \text{atp}_{\mathcal{F}}(\bar{\mathcal{G}}_1)} e.$$

In the inductive step of the construction we define

$$\begin{aligned} e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k+1} := & ((\text{paths}_{k+1}^{\mathcal{F}} \cap \varphi_{\bar{\mathcal{G}}_1, \text{posatoms}}^{\mathcal{F}}) - \varphi_{\bar{\mathcal{G}}_1, \text{negatoms}}^{\mathcal{F}}) \\ & \cap \varphi_{\bar{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{F},k+1} \cap \varphi_{\bar{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F},k+1} \\ & \cap \varphi_{\bar{\mathcal{G}}_1, \text{projection forth}}^{\mathcal{F},k+1} \cap \varphi_{\bar{\mathcal{G}}_1, \text{projection back}}^{\mathcal{F},k+1} \end{aligned}$$

where (with  $V_1$  the node set of  $\mathcal{G}_1$ )

$$\varphi_{\bar{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{F},k+1} := \bigcap_{\substack{c_1 \in V_1 \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \\ (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},k} \circ e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F},k};$$



$$\begin{aligned} \varphi_{\bar{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F}, k+1} := & \text{paths}_{k+1}^{\mathcal{F}} - \\ & \bigcup_{V \subseteq V_1} \left( \bigcap_{\substack{c_1 \in V \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \\ (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} (\text{paths}_k^{\mathcal{F}} - e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, k}) \right. \\ & \left. \circ \bigcap_{\substack{c_1 \in V_1 - V \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \\ (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} (\text{paths}_k^{\mathcal{F}} - e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F}, k}) \right); \end{aligned}$$

and the Projection Forth and Projection Back expressions are omitted from  $e_{\bar{\mathcal{G}}_1}^{\mathcal{F}, k}$  if  $a_1 \neq b_1$  or  $\pi$  is not present in  $(\mathcal{F})$ ; otherwise they are defined as

$$\begin{aligned} \varphi_{\bar{\mathcal{G}}_1, \text{projection forth}}^{\mathcal{F}, k+1} := & \bigcap_{\substack{c_1 \in V_1 \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} \pi_1(e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, k}) \\ & \cap \bigcap_{\substack{c_1 \in V_1 \\ (c_1, a_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} \pi_2(e_{(\mathcal{G}_1, c_1, a_1)}^{\mathcal{F}, k}) \end{aligned}$$

and

$$\begin{aligned} \varphi_{\bar{\mathcal{G}}_1, \text{projection back}}^{\mathcal{F}, k+1} := & (1' - \pi_1(\text{paths}_k^{\mathcal{F}} - \bigcup_{\substack{c_1 \in V_1 \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, k})) \\ & - \pi_2(\text{paths}_k^{\mathcal{F}} - \bigcup_{\substack{c_1 \in V_1 \\ (c_1, a_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} e_{(\mathcal{G}_1, c_1, a_1)}^{\mathcal{F}, k}). \end{aligned}$$

To show the correctness of the above expression, first note that  $e_{\bar{\mathcal{G}}_1}^{\mathcal{F}, k}$  indeed belongs to  $\mathcal{F}_k$ . It remains to be shown that, for any structure  $\mathcal{G}_2$ ,

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{F}, k}(\mathcal{G}_2) = \{(a_2, b_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \mid \bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)\}.$$

This can be shown by induction on  $k$ . For  $k = 0$ , it is clear that  $(a_2, b_2) \in e_{\bar{\mathcal{G}}_1}^{\mathcal{F}, 0}(\mathcal{G}_2)$  iff  $\text{atp}_{\mathcal{F}}(\bar{\mathcal{G}}_1) = \text{atp}_{\mathcal{F}}(\mathcal{G}_2)$ , i.e., iff the Atoms condition for  $\bar{\mathcal{G}}_1 \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)$  is satisfied. Furthermore, assuming the induction hypothesis for  $k$ , the following equivalences are readily verified for any  $(a_2, b_2) \in \text{paths}_{k+1}^{\mathcal{F}}(\mathcal{G}_2)$ : (the second and third equivalences assume  $a_1 = b_1$ )

- $(a_2, b_2) \in \varphi_{\bar{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{F}, k+1}(\mathcal{G}_2)$  iff the Composition Forth condition for  $\bar{\mathcal{G}}_1 \simeq_{k+1}^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)$  is satisfied.
- $(a_2, b_2) \in ((\varphi_{\bar{\mathcal{G}}_1, \text{posatoms}}^{\mathcal{F}} - \varphi_{\bar{\mathcal{G}}_1, \text{negatoms}}^{\mathcal{F}}) \cap \varphi_{\bar{\mathcal{G}}_1, \text{projection forth}}^{\mathcal{F}, k+1})(\mathcal{G}_2)$  iff the Atoms and Projection Forth conditions for  $\bar{\mathcal{G}}_1 \simeq_{k+1}^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)$  are satisfied.

- $(a_2, b_2) \in ((\varphi_{\bar{\mathcal{G}}_1, \text{posatoms}}^{\mathcal{F}} - \varphi_{\bar{\mathcal{G}}_1, \text{negatoms}}^{\mathcal{F}}) \cap \varphi_{\bar{\mathcal{G}}_1, \text{projection back}}^{\mathcal{F}, k+1})(\mathcal{G}_2)$  iff the Atoms and Projection Back conditions for  $\bar{\mathcal{G}}_1 \simeq_{k+1}^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)$  are satisfied.

It remains to prove that  $(a_2, b_2) \in \varphi_{\bar{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F}, k+1}(\mathcal{G}_2)$  iff the Composition Back condition for  $\bar{\mathcal{G}}_1 \simeq_{k+1}^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)$  is satisfied. By inspection of the expression, we see that  $(a_2, b_2) \in \varphi_{\bar{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F}, k+1}(\mathcal{G}_2)$  iff there is no subset  $V \subseteq V_1$  for which there exists  $c_2$  in  $V_2$  such that

$$\begin{aligned} \forall c_1 \in V : ((a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \wedge (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)) \Rightarrow \\ ((a_2, c_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \wedge (a_2, c_2) \notin e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, k}(\mathcal{G}_2)) \end{aligned}$$

and

$$\begin{aligned} \forall c_1 \notin V : ((a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \wedge (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)) \Rightarrow \\ ((c_2, b_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \wedge (c_2, b_2) \notin e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F}, k}(\mathcal{G}_2)). \end{aligned}$$

In other words, and by induction, for every  $V \subseteq V_1$  and for all  $c_2$  in  $V_2$ , we have

$$\begin{aligned} \exists c_1 \in V : (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \wedge (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \wedge \\ ((a_2, c_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \Rightarrow (\mathcal{G}_1, a_1, c_1) \simeq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)) \end{aligned}$$

or

$$\begin{aligned} \exists c_1 \notin V : (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \wedge (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \wedge \\ ((c_2, b_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \Rightarrow (\mathcal{G}_1, c_1, b_1) \simeq_k^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)). \end{aligned}$$

More formally, for every  $c_2$  in  $V_2$ , we have

$$\begin{aligned} \bigwedge_{V \subseteq V_1} \left( \bigvee_{\substack{c_1 \in V \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \\ (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} ((a_2, c_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \Rightarrow (\mathcal{G}_1, a_1, c_1) \simeq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)) \right. \\ \left. \vee \bigvee_{\substack{c_1 \in V_1 - V \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \\ (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} ((c_2, b_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \Rightarrow (\mathcal{G}_1, c_1, b_1) \simeq_k^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2))) \right). \end{aligned}$$

Now, using commutativity of the logical ‘or’, and distributivity of the logical ‘and’ over the logical ‘or’, we can equivalently write the above as follows, still for every  $c_2$  in  $V_2$ :

$$\begin{aligned} \bigvee_{\substack{c_1 \in V_1 \\ (a_1, c_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1) \\ (c_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)}} ((a_2, c_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \Rightarrow (\mathcal{G}_1, a_1, c_1) \simeq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)) \\ \wedge ((c_2, b_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \Rightarrow (\mathcal{G}_1, c_1, b_1) \simeq_k^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)), \end{aligned}$$

which is exactly the Composition Back condition.  $\square$

We conclude:

**Theorem 15.**  $\bar{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \bar{\mathcal{G}}_2$  if and only if  $\bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$ .

*Proof.* The if-direction is given by the Invariance Lemma. For the only-if-direction, we use the Representation Lemma. Let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\bar{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$ . If  $(a_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$ , then, since trivially  $\bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \bar{\mathcal{G}}_1$ , we have  $(a_1, b_1) \in e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_1)$ , whence  $(a_2, b_2) \in e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_2)$ , whence  $\bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$  as desired. If  $(a_1, b_1) \notin \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$ , then the Composition Forth and Back and the Projection Forth and Back conditions are void. The atoms condition is satisfied because  $\bar{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \bar{\mathcal{G}}_2$ . We again conclude that  $\bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$ .  $\square$

## 4.2 Fragments with the residuals

In this section,  $\mathcal{F}$  is an arbitrary fixed fragment of the calculus of relations in which both set difference or complementation, and the residual operations are present.

**Definition 16** (Bisimilarity incl. residuals). Let  $\mathcal{F}$  be a fragment of the calculus of relations containing the difference and residual operations. Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\bar{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be marked structures with node sets  $V_1$  and  $V_2$ , respectively. We say that  $\bar{\mathcal{G}}_1$  is  $\mathcal{F}$ -bisimilar to  $\bar{\mathcal{G}}_2$  up to depth  $k$ , denoted  $\bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$ , if the following conditions are satisfied:

**Atoms**  $\text{atp}_{\mathcal{F}}(\bar{\mathcal{G}}_1) = \text{atp}_{\mathcal{F}}(\bar{\mathcal{G}}_2)$ ;

**Composition Forth** if  $k = 1$ , then, for every  $c_1$  in  $V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_1)$ , there exists  $c_2$  in  $V_2$  such that both  $(\mathcal{G}_1, a_1, c_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

if  $k > 1$ , then, for every  $c_1$  in  $V_1$ , there exists  $c_2$  in  $V_2$  such that both  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

**Projection Forth** if  $\pi$  is in  $(\mathcal{F})$ , if  $k = 1$ , and if  $a_1 = b_1$ , then, for every  $c_1$  in  $V_1$  with  $(a_1, c_1)$  in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_1)$  (resp.,  $(c_1, a_1)$  in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_1)$ ), there exists  $c_2$  in  $V_2$  such that  $(\mathcal{G}_1, a_1, c_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  (resp.,  $(\mathcal{G}_1, c_1, a_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ );

if  $\pi$  is in  $(\mathcal{F})$ , if  $k > 1$ , and if  $a_1 = b_1$ , then, for every  $c_1$  in  $V_1$ , there exists  $c_2$  in  $V_2$  such that  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  (resp.,  $(\mathcal{G}_1, c_1, a_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ );

**Left Residual Forth** if  $k = 1$ , then, for every  $c_2$  in  $V_2$  with  $(b_2, c_2)$  in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_2)$ , there exists  $c_1$  in  $V_1$  such that both  $(\mathcal{G}_2, b_2, c_2) \simeq_0^{\mathcal{F}} (\mathcal{G}_1, b_1, c_1)$  and  $(\mathcal{G}_1, a_1, c_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$ ;

if  $k > 1$ , then, for every  $c_2$  in  $V_2$ , there exists  $c_1$  in  $V_1$  such that both  $(\mathcal{G}_2, b_2, c_2) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, b_1, c_1)$  and  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$ ;

**Right Residual Forth** if  $k = 1$ , then, for every  $c_2$  in  $V_2$  with  $(c_2, a_2)$  in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_2)$ , there exists  $c_1$  in  $V_1$  such that both  $(\mathcal{G}_2, c_2, a_2) \simeq_0^{\mathcal{F}} (\mathcal{G}_1, c_1, a_1)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

if  $k > 1$ , then, for every  $c_2$  in  $V_2$ , there exists  $c_1$  in  $V_1$  such that both  $(\mathcal{G}_2, c_2, a_2) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, c_1, a_1)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

**Composition Back** if  $k = 1$ , then, for every  $c_2$  in  $V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_2)$ , there exists  $c_1$  in  $V_1$  such that both  $(\mathcal{G}_1, a_1, c_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

if  $k > 1$ , then, for every  $c_2$  in  $V_2$ , there exists  $c_1$  in  $V_1$  such that both  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

**Projection Back** if  $\pi$  is in  $(\mathcal{F})$ , if  $k = 1$ , and if  $a_1 = b_1$ , then, for every  $c_2$  in  $V_2$  with  $(a_2, c_2)$  in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_2)$  (resp.,  $(c_2, a_2)$  in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_2)$ ), there exists  $c_1$  in  $V_1$  such that  $(\mathcal{G}_1, a_1, c_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  (resp.,  $(\mathcal{G}_1, c_1, a_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ );

if  $\pi$  is in  $(\mathcal{F})$ , if  $k > 1$ , and if  $a_1 = b_1$ , then, for every  $c_2$  in  $V_2$ , there exists  $c_1$  in  $V_1$  such that  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  (resp.,  $(\mathcal{G}_1, c_1, a_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ );

**Left Residual Back** if  $k = 1$ , then, for every  $c_1$  in  $V_1$ , there exists  $c_2$  in  $V_2$  such that  $(\mathcal{G}_2, b_2, c_2) \simeq_0^{\mathcal{F}} (\mathcal{G}_1, b_1, c_1)$ , and either  $(a_2, c_2) \notin \text{paths}_0^{\mathcal{F}}(\mathcal{G}_2)$ , or  $(\mathcal{G}_1, a_1, c_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$ ;

if  $k > 1$ , then, for every  $c_1$  in  $V_1$ , there exists  $c_2$  in  $V_2$  such that both  $(\mathcal{G}_1, b_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, b_2, c_2)$  and  $(\mathcal{G}_2, a_2, c_2) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, a_1, c_1)$ ;

**Right Residual Back** if  $k = 1$ , then, for every  $c_1$  in  $V_1$ , there exists  $c_2$  in  $V_2$  such that  $(\mathcal{G}_1, c_1, a_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ , and either  $(c_2, b_2) \notin \text{paths}_0^{\mathcal{F}}(\mathcal{G}_2)$ , or  $(\mathcal{G}_2, c_2, b_2) \simeq_0^{\mathcal{F}} (\mathcal{G}_1, c_1, b_1)$ ;

if  $k > 1$ , then, for every  $c_1$  in  $V_1$ , there exists  $c_2$  in  $V_2$  such that both  $(\mathcal{G}_1, c_1, a_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$  and  $(\mathcal{G}_2, c_2, b_2) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, c_1, b_1)$ .

Note that the conditions associated to the residuals have a special case for  $k = 1$ . This is because the operands of a residual in an expression of degree  $k = 1$ , being expressions of degree 0, are necessarily contained in  $\text{paths}_0$ , whereas this need no longer be the case in higher-degree expressions; recall Lemma 3.

**Lemma 17** (Invariance lemma—bisimilarity incl. residuals). *If  $\overline{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ , then  $\overline{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \overline{\mathcal{G}}_2$ .*

*Proof.* The proof proceeds like the proof of Lemma 13; we only discuss what is new.

For the case where  $e$  is  $e_1 \circ e_2$ , consider the only-if, i.e., assume  $(a_1, b_1) \in e(\mathcal{G}_1)$ . By definition of composition there exists  $c_1$  in  $V_1$  with  $(a_1, c_1) \in e_1(\mathcal{G}_1)$  and  $(c_1, b_1) \in e_2(\mathcal{G}_1)$ . If  $k = 1$ , we have that  $e_1$  and  $e_2$  are in  $\mathcal{F}_0$ . By Lemma 3, we have that both  $(a_1, c_1)$  and  $(c_1, b_1)$  are in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_1)$ . By the Composition Forth condition in the definition of bisimilarity, there exists  $c_2$  in  $V_2$  such that both  $(\mathcal{G}_1, a_1, c_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \simeq_0^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ . By induction, we have  $(a_2, c_2) \in e_1(\mathcal{G}_2)$  and  $(c_2, b_2) \in e_2(\mathcal{G}_2)$ , and hence  $(a_2, b_2) \in e_1 \circ e_2(\mathcal{G}_2)$ . If  $k > 1$ , the argument is again the same as in the proof of Lemma 13.

For the case where  $e$  is  $e_1 / e_2$ , consider the only-if, i.e., assume  $(a_1, b_1) \in e(\mathcal{G}_1)$ . Suppose now that  $(a_2, b_2) \notin e_1 / e_2(\mathcal{G}_2)$ . Then, by definition of the left residual, there exists  $c_2$  in  $V_2$  such that (1)  $(b_2, c_2) \in e_2(\mathcal{G}_2)$ , and (2)  $(a_2, c_2) \notin e_1(\mathcal{G}_2)$ . If  $k = 1$ , we have that  $e_2$  is in  $\mathcal{F}_0$ . By Lemma 3, we have that  $(b_2, c_2)$  is in  $\text{paths}_0^{\mathcal{F}}(\mathcal{G}_2)$ . By the Left Residual Forth condition in the definition of bisimilarity, there exists  $c_1$  in  $V_1$  such that (1)  $(\mathcal{G}_2, b_2, c_2) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, b_1, c_1)$ , and (2)  $(\mathcal{G}_1, a_1, c_1) \simeq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$ . By the induction hypothesis, we obtain  $(b_1, c_1) \in e_2(\mathcal{G}_1)$  and  $(a_1, c_1) \notin e_1(\mathcal{G}_1)$ . Now, this  $c_1$  contradicts that  $(a_1, b_1) \in e_1 / e_2(\mathcal{G}_1)$ . If  $k > 1$ , the argument is similar. Also, the argument for the if-direction is similar to the argument for the only-if direction; it uses the Left Residual Back condition in the definition of bisimilarity.

The case of a right residual is analogous to that of a left residual.  $\square$

**Lemma 18** (Representation lemma—bisimilarity incl. residuals). *Let  $\mathcal{F}$  be a fragment of the calculus of relations containing the difference and the residual operations. Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  be a marked structure. There exists an expression  $e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}$  in  $\mathcal{F}_k$  such that for every structure  $\mathcal{G}_2$  with node set  $V_2$ :*

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{F},0}(\mathcal{G}_2) = \{(a_2, b_2) \in \text{paths}_0^{\mathcal{F}}(\mathcal{G}_2) \mid \bar{\mathcal{G}}_1 \simeq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)\}, \text{ and}$$

$$\text{for } k \geq 1: e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_2) = \{(a_2, b_2) \in V_2 \times V_2 \mid \bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)\}.$$

*Proof.* The proof proceeds like that of Lemma 14; we only discuss what is new. The expression for  $k = 0$  is the same as before. In the inductive step, the expression has the following general form:

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k+1} := ((1 \cap \varphi_{\bar{\mathcal{G}}_1, \text{posatoms}}^{\mathcal{F}}) - \varphi_{\bar{\mathcal{G}}_1, \text{negatoms}}^{\mathcal{F}}) \\
\cap \varphi_{\bar{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{F},k+1} \cap \varphi_{\bar{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F},k+1} \\
\cap \varphi_{\bar{\mathcal{G}}_1, \text{projection forth}}^{\mathcal{F},k+1} \cap \varphi_{\bar{\mathcal{G}}_1, \text{projection back}}^{\mathcal{F},k+1} \\
\cap \psi_{\bar{\mathcal{G}}_1, \text{leftres forth}}^{\mathcal{F},k+1} \cap \psi_{\bar{\mathcal{G}}_1, \text{leftres back}}^{\mathcal{F},k+1} \\
\cap \psi_{\bar{\mathcal{G}}_1, \text{rightres forth}}^{\mathcal{F},k+1} \cap \psi_{\bar{\mathcal{G}}_1, \text{rightres back}}^{\mathcal{F},k+1}.$$

The subexpressions have different definitions depending on whether the degree is 1 or higher. For degree 1, the subexpressions for composition and projection (back and forth) are as in Lemma 14. The other subexpressions for degree 1 are

as follows:

$$\begin{aligned}
\psi_{\bar{\mathcal{G}}_1, \text{leftres forth}}^{\mathcal{F},1} &:= \bigcap_{V \subseteq V_1} \left[ \left( \bigcup_{c_1 \in V_1 - V} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},0} \right) / \left( \text{paths}_0^{\mathcal{F}} - \bigcup_{c_1 \in V} e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F},0} \right) \right]; \\
\psi_{\bar{\mathcal{G}}_1, \text{leftres back}}^{\mathcal{F},1} &:= \left( \bigcup_{c_1 \in V_1} \left( \text{paths}_0^{\mathcal{F}} - e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},0} \right) / e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F},0} \right)^c; \\
\psi_{\bar{\mathcal{G}}_1, \text{rightres forth}}^{\mathcal{F},1} &:= \bigcap_{V \subseteq V_1} \left[ \left( \text{paths}_0^{\mathcal{F}} - \bigcup_{c_1 \in V} e_{(\mathcal{G}_1, c_1, a_1)}^{\mathcal{F},0} \right) \setminus \left( \bigcup_{c_1 \in V_1 - V} e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F},0} \right) \right]; \\
\psi_{\bar{\mathcal{G}}_1, \text{rightres back}}^{\mathcal{F},1} &:= \left( \bigcup_{c_1 \in V_1} e_{(\mathcal{G}_1, c_1, a_1)}^{\mathcal{F},0} \setminus \left( \text{paths}_0^{\mathcal{F}} - e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F},0} \right) \right)^c.
\end{aligned}$$

For  $k > 0$ , we define

$$\begin{aligned}
\varphi_{\bar{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{F},k+1} &:= \bigcap_{c_1 \in V_1} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},k} \circ e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F},k}; \\
\varphi_{\bar{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F},k+1} &:= \left( \bigcup_{V \subseteq V_1} \left( \bigcap_{c_1 \in V} (e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},k})^c \circ \bigcap_{c_1 \in V_1 - V} (e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F},k})^c \right) \right)^c; \\
\varphi_{\bar{\mathcal{G}}_1, \text{projection forth}}^{\mathcal{F},k+1} &:= \bigcap_{c_1 \in V_1} \pi_1(e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},k}) \cap \bigcap_{c_1 \in V_1} \pi_2(e_{(\mathcal{G}_1, c_1, a_1)}^{\mathcal{F},k}); \\
\varphi_{\bar{\mathcal{G}}_1, \text{projection back}}^{\mathcal{F},k+1} &:= \left( \pi_1 \left( \bigcup_{c_1 \in V_1} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},k} \right)^c - \pi_2 \left( \bigcup_{c_1 \in V_1} e_{(\mathcal{G}_1, c_1, a_1)}^{\mathcal{F},k} \right)^c \right)^c; \\
\psi_{\bar{\mathcal{G}}_1, \text{leftres forth}}^{\mathcal{F},k+1} &:= \bigcap_{V \subseteq V_1} \left[ \left( \bigcup_{c_1 \in V_1 - V} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},k} \right) / \left( \bigcup_{c_1 \in V} e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F},k} \right) \right]^c; \\
\psi_{\bar{\mathcal{G}}_1, \text{leftres back}}^{\mathcal{F},k+1} &:= \left( \bigcup_{c_1 \in V_1} \left( e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},k} \right)^c / e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F},k} \right)^c; \\
\psi_{\bar{\mathcal{G}}_1, \text{rightres forth}}^{\mathcal{F},k+1} &:= \bigcap_{V \subseteq V_1} \left[ \left( \bigcup_{c_1 \in V} e_{(\mathcal{G}_1, c_1, a_1)}^{\mathcal{F},k} \right)^c \setminus \left( \bigcup_{c_1 \in V_1 - V} e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F},k} \right) \right]^c; \\
\psi_{\bar{\mathcal{G}}_1, \text{rightres back}}^{\mathcal{F},k+1} &:= \left( \bigcup_{c_1 \in V_1} e_{(\mathcal{G}_1, c_1, a_1)}^{\mathcal{F},k} \setminus \left( e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F},k} \right)^c \right)^c.
\end{aligned}$$

Note that the above expressions use complementation, which is allowed since complementation is definable as  $e^c = (0/0) - e$ , an expression which has the same degree as  $e$  provided the degree of  $e$  is at least one. (Recall that  $0/0 \equiv 1$ .)

To show correctness, we show that  $(a_2, b_2) \in \psi_{\bar{\mathcal{G}}_1, \text{leftres forth}}^{\mathcal{F},1}(\mathcal{G}_2)$  iff the Left Residual Forth condition for  $\bar{\mathcal{G}}_1 \simeq_1^{\mathcal{F}}(\mathcal{G}_2, a_2, b_2)$  is satisfied. Inspecting the expression, we see that  $(a_2, b_2) \in \psi_{\bar{\mathcal{G}}_1, \text{leftres forth}}^{\mathcal{F},1}(\mathcal{G}_2)$  iff for all subsets  $V \subseteq V_1$  and for all  $c_2 \in V_2$ , if  $(b_2, c_2)$  is in  $\text{paths}_0(\mathcal{G}_2) - \bigcup_{c_1 \in V} e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F},0}(\mathcal{G}_2)$ , then  $(a_2, c_2)$  is

in  $\bigcup_{c_1 \in V_1 - V} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2)$ . In other words, for all subsets  $V \subseteq V_1$  and for all  $c_2 \in V_2$  with  $(b_2, c_2) \in \text{paths}_0(\mathcal{G}_2)$ :

$$(b_2, c_2) \in \bigcup_{c_1 \in V} e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2) \quad \vee \quad (a_2, c_2) \in \bigcup_{c_1 \in V_1 - V} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2).$$

Equivalently, for all  $c_2 \in V_2$  with  $(b_2, c_2) \in \text{paths}_0(\mathcal{G}_2)$ :

$$\bigwedge_{V \subseteq V_1} \left( \bigvee_{c_1 \in V} (b_2, c_2) \in e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2) \quad \vee \quad \bigvee_{c_1 \in V_1 - V} (a_2, c_2) \in e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2) \right).$$

Now, using commutativity of the logical ‘or’, and distributivity of the logical ‘and’ over the logical ‘or’, we can equivalently write this as follows: for all  $c_2 \in V_2$  with  $(b_2, c_2) \in \text{paths}_0(\mathcal{G}_2)$

$$\bigvee_{c_1 \in V_1} \left( (b_2, c_2) \in e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2) \quad \wedge \quad (a_2, c_2) \in e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2) \right),$$

which is, by the result for  $k = 0$ , exactly the Left Residual Forth condition.

We next show that  $(a_2, b_2) \in \psi_{\mathcal{G}_1, \text{leftres back}}^{\mathcal{F}, 1}$  iff the Left Residual Back condition for  $\bar{\mathcal{G}}_1 \simeq_1^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)$  is satisfied. Inspecting the expression, we see that  $(a_2, b_2) \in \psi_{\mathcal{G}_1, \text{leftres back}}^{\mathcal{F}, 1}$  iff for every  $c_1 \in V_1$  we have

$$(a_2, b_2) \notin \left( \text{paths}_0^{\mathcal{F}} - e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, 0} \right) / e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2).$$

By definition of the left residual, the above means that there exists  $c_2 \in V_2$  such that

$$(b_2, c_2) \in e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2) \quad \wedge \quad (a_2, c_2) \notin \text{paths}_0^{\mathcal{F}} - e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2).$$

Equivalently, for each  $c_1 \in V_1$  there exists  $c_2 \in V_2$  such that

$$(b_2, c_2) \in e_{(\mathcal{G}_1, b_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2) \quad \wedge \quad \left( (a_2, c_2) \notin \text{paths}_0^{\mathcal{F}}(\mathcal{G}_2) \vee e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F}, 0}(\mathcal{G}_2) \right),$$

which is, by the result for  $k = 0$ , exactly the Left Residual Back condition.

Similar arguments are used to show the corresponding equivalences regarding  $\psi_{\mathcal{G}_1, \text{leftres forth}}^{\mathcal{F}, k+1}$  and  $\psi_{\mathcal{G}_1, \text{leftres back}}^{\mathcal{F}, k+1}$ , and also the arguments for the right residual are similar.  $\square$

The results in Section 4.1 and Section 4.2 lead us to

**Theorem 19.** *Let  $\mathcal{F}$  be a fragment of the calculus of relations containing the difference operation. Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\bar{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be marked structures. Then,  $\bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$  if and only if  $\bar{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \bar{\mathcal{G}}_2$ .*

*Proof.* The only-if direction has already been given by Invariance Lemma 13 (for fragments not containing residual operations) and Invariance Lemma 17 (for fragments that do contain the residuals).

The if-direction follows from Representation Lemma 14 (for fragments not containing residual operations) and from Representation Lemma 18 (for fragments that do contain a residual operation). In particular, let  $\mathcal{F}$  be a fragment not containing residual operations and assume  $\bar{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \bar{\mathcal{G}}_2$ . We consider two cases: (1)  $(a_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$ , and (2)  $(a_1, b_1) \notin \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$ . For case (1), by Representation Lemma 14, we have  $(a_1, b_1) \in e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_1)$  and therefore also  $(a_2, b_2) \in e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_2)$ . By definition of  $e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}$ , we obtain  $\bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$ . For case (2), the bisimulation conditions are vacuously true.

For the case where  $\mathcal{F}$  is a fragment that does contain the residuals, assume again  $\bar{\mathcal{G}}_1 \equiv_{\mathcal{F}_k} \bar{\mathcal{G}}_2$ . We again consider two cases. If  $k \geq 1$ , we use Representation Lemma 18 to obtain  $\bar{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$ . If  $k = 0$ , it is clear that  $\bar{\mathcal{G}}_1 \equiv_{\mathcal{F}_0} \bar{\mathcal{G}}_2$  implies the atoms condition of  $\bar{\mathcal{G}}_1 \simeq_0^{\mathcal{F}} \bar{\mathcal{G}}_2$ .  $\square$

## 5 Similarity for fragments without set difference

### 5.1 Fragments not containing residual operations

In Section 3.1, we defined similarity for the fragment  $\mathcal{C}(\bar{\pi})$  (Definition 7), and showed the invariance and representation lemmas (Lemma 9 and Lemma 10). For completeness, we list the more general definition of similarity for fragments not containing residual operations and give, without proof, the corresponding invariance lemma and representation lemma.

**Definition 20** (Similarity excl. residuals). Let  $\mathcal{F}$  be a fragment of the calculus of relations not containing the difference operation and not containing residual operations. Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\bar{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be marked structures with node sets  $V_1$  and  $V_2$ , respectively. We say that  $\bar{\mathcal{G}}_1$  is  $\mathcal{F}$ -similar to  $\bar{\mathcal{G}}_2$  up to depth  $k$ , denoted  $\bar{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$ , if the following conditions are satisfied:

**Atoms**  $\text{atp}_{\mathcal{F}}(\bar{\mathcal{G}}_1) \subseteq \text{atp}_{\mathcal{F}}(\bar{\mathcal{G}}_2)$ ;

**Composition Forth** if  $k > 0$ , then, for every  $c_1$  in  $V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ , there exists  $c_2$  in  $V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$  such that both  $(\mathcal{G}_1, a_1, c_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

**Projection Forth** if  $\pi$  is in  $(\mathcal{F})$ , if  $k > 0$ , and if  $a_1 = b_1$ , then, for every  $c_1$  in  $V_1$  with  $(a_1, c_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$  (resp.,  $(c_1, a_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ ), there exists  $c_2$  in  $V_2$  with  $(a_2, c_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$  (resp.,  $(c_2, a_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$ ) such that  $(\mathcal{G}_1, a_1, c_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  (resp.,  $(\mathcal{G}_1, c_1, a_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ );

**Coprojection forth** if  $\bar{\pi}$  is in  $(\mathcal{F})$ , if  $k > 0$ , and if  $a_1 = b_1$ , then, for every  $c_2$  in  $V_2$  with  $(a_2, c_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$  (resp.,  $(c_2, a_2)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$ ), there



exists  $c_1$  in  $V_1$  with  $(a_1, c_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$  (resp.,  $(c_1, a_1)$  in  $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ ) such that  $(\mathcal{G}_2, a_2, c_2) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, a_1, c_1)$  (resp.,  $(\mathcal{G}_2, c_2, a_2) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, c_1, a_1)$ ).

**Lemma 21** (Invariance lemma—similarity excl. residuals). *Let  $\mathcal{F}$  be a fragment of the calculus of relations not containing the difference operation and not containing residual operations. Let  $k$  be a natural number and let  $\overline{\mathcal{G}}_1$  and  $\overline{\mathcal{G}}_2$  be marked structures. If  $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ , then  $\overline{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}_k} \overline{\mathcal{G}}_2$ .*

**Lemma 22** (Representation lemma—similarity excl. residuals). *Let  $\mathcal{F}$  be a fragment of the calculus of relations not containing the difference operation and not containing residual operations. Let  $k$  be a natural number and let  $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  be a marked structure. There exists an expression  $e_{\overline{\mathcal{G}}_1}^{\mathcal{F}, k}$  in  $\mathcal{F}_k$  such that for every structure  $\mathcal{G}_2$ :*

$$e_{\overline{\mathcal{G}}_1}^{\mathcal{F}, k}(\mathcal{G}_2) = \{(a_2, b_2) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \mid \overline{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)\}.$$

## 5.2 Fragments with the residuals

**Definition 23** (Similarity incl. residuals). Let  $\mathcal{F}$  be a fragment of the calculus of relations containing a residual operation and not containing the difference operation. Let  $k$  be a natural number and let  $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be marked structures with node sets  $V_1$  and  $V_2$ , respectively. We say that  $\overline{\mathcal{G}}_1$  is  $\mathcal{F}$ -similar to  $\overline{\mathcal{G}}_2$  up to depth  $k$ , denoted  $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ , if the following conditions are satisfied:

**Atoms**  $\text{atp}_{\mathcal{F}}(\overline{\mathcal{G}}_1) \subseteq \text{atp}_{\mathcal{F}}(\overline{\mathcal{G}}_2)$ ;

**Composition Forth** if  $k > 0$ , then, for every  $c_1$  in  $V_1$ , there exists  $c_2$  in  $V_2$  such that both  $(\mathcal{G}_1, a_1, c_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ ;

**Projection Forth** if  $\pi$  is in  $(\mathcal{F})$ , if  $k > 0$ , and if  $a_1 = b_1$ , then, for every  $c_1$  in  $V_1$ , there exists  $c_2$  in  $V_2$  such that  $(\mathcal{G}_1, a_1, c_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$  (resp.,  $(\mathcal{G}_1, c_1, a_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, a_2)$ );

**Left Residual Forth** if  $k > 0$ , then, for every  $c_2$  in  $V_2$ , there exists  $c_1$  in  $V_1$  such that both  $(\mathcal{G}_2, b_2, c_2) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, b_1, c_1)$  and  $(\mathcal{G}_1, a_1, c_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, a_2, c_2)$ ;

**Right Residual Forth** if  $k > 0$ , then, for every  $c_2$  in  $V_2$ , there exists  $c_1$  in  $V_1$  such that both  $(\mathcal{G}_2, c_2, a_2) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_1, c_1, a_1)$  and  $(\mathcal{G}_1, c_1, b_1) \preceq_{k-1}^{\mathcal{F}} (\mathcal{G}_2, c_2, b_2)$ .

Unlike the situation for bisimilarity, here no special case for  $k = 1$  is needed, due to the general principle that whenever  $(a_1, c_1) \notin \text{paths}_0^{\mathcal{F}}(\mathcal{G}_1)$  (or  $(c_1, b_1) \notin \text{paths}_0^{\mathcal{F}}(\mathcal{G}_1)$ ) then  $(\mathcal{G}_1, a_1, c_1) \preceq_0^{\mathcal{F}} \overline{\mathcal{G}}_2$  (or  $(\mathcal{G}_1, c_1, b_1) \preceq_0^{\mathcal{F}} \overline{\mathcal{G}}_2$ ) holds trivially for any  $\overline{\mathcal{G}}_2$ .

The following lemma is proven in the same way as the preceding invariance lemmas.

**Lemma 24** (Invariance lemma—similarity incl. residuals). *Let  $\mathcal{F}$  be a fragment of the calculus of relations containing a residual operation and not containing the difference operation. Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1$  and  $\bar{\mathcal{G}}_2$  be marked structures. If  $\bar{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$ , then  $\bar{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}_k} \bar{\mathcal{G}}_2$ .*

*Proof.* The proof proceeds as the proof of Lemma 17.  $\square$

We finally show

**Lemma 25** (Representation lemma—similarity incl. residuals). *Let  $\mathcal{F}$  be a fragment of the calculus of relations containing a residual operation and not containing the difference operation. Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  be a marked structure. There exists an expression  $e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}$  in  $\mathcal{F}_k$  such that for every structure  $\mathcal{G}_2$ :*

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{F},0}(\mathcal{G}_2) = \{(a_2, b_2) \in \text{paths}_0^{\mathcal{F}}(\mathcal{G}_2) \mid \bar{\mathcal{G}}_1 \preceq_0^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)\}, \text{ and}$$

$$\text{for } k \geq 1: e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_2) = \{(a_2, b_2) \in V_2 \times V_2 \mid \bar{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)\}.$$

*Proof.* The proof proceeds as for the preceding Representation Lemmas 10 and 18. The expression for  $k = 0$  is the same as in Lemma 10. In the inductive step the expression has the following general form:

$$e_{\bar{\mathcal{G}}_1}^{\mathcal{F},k+1} := 1 \cap \chi_{\bar{\mathcal{G}}_1, \text{atoms}}^{\mathcal{F}} \cap \psi_{\bar{\mathcal{G}}_1, \text{composition forth}}^{\mathcal{F},k+1} \cap \psi_{\bar{\mathcal{G}}_1, \text{projection forth}}^{\mathcal{F},k+1} \cap \xi_{\bar{\mathcal{G}}_1, \text{leftres forth}}^{\mathcal{F},k+1} \cap \xi_{\bar{\mathcal{G}}_1, \text{rightres forth}}^{\mathcal{F},k+1}$$

with

$$\xi_{\bar{\mathcal{G}}_1, \text{leftres forth}}^{\mathcal{F},k+1} := \bigcap_{V \subseteq V_1} \left[ \left( \bigcup_{c_1 \in V_1 - V} e_{(\mathcal{G}_1, a_1, c_1)}^{\mathcal{F},k} \right) / \left( \bigcap_{c_1 \in V} \bigcup_{\substack{e \in \mathcal{F}_k \\ (b_1, c_1) \notin e(\mathcal{G}_1)}} e \right) \right]$$

and

$$\xi_{\bar{\mathcal{G}}_1, \text{rightres forth}}^{\mathcal{F},k+1} := \bigcap_{V \subseteq V_1} \left[ \left( \bigcap_{c_1 \in V} \bigcup_{\substack{e \in \mathcal{F}_k \\ (c_1, a_1) \notin e(\mathcal{G}_1)}} e \right) \setminus \left( \bigcup_{c_1 \in V_1 - V} e_{(\mathcal{G}_1, c_1, b_1)}^{\mathcal{F},k} \right) \right].$$

The correctness argument involves no new insights beyond the proofs of the previous representation lemmas.  $\square$

### 5.3 Characterization theorem

From the preceding two subsections, in a similar way as Theorem 19, we obtain

**Theorem 26.** *Let  $\mathcal{F}$  be a fragment of the calculus of relations not containing the difference operation. Let  $k$  be a natural number and let  $\bar{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$  and  $\bar{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$  be marked structures. Then,  $\bar{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} \bar{\mathcal{G}}_2$  if and only if  $\bar{\mathcal{G}}_1 \Rightarrow_{\mathcal{F}_k} \bar{\mathcal{G}}_2$ .*

## 6 Indistinguishability of finite structures

The characterizations we have given of when two structures are indistinguishable by expressions of  $\mathcal{F}_k$ , for some fixed degree  $k$  and some fixed fragment  $\mathcal{F}$ , are valid for arbitrary structures. We may now ask, given two *finite* structures as input, if it is effectively decidable whether or not they are indistinguishable by expressions of  $\mathcal{F}_k$ . Since there are only a fixed, finite number of expressions in  $\mathcal{F}_k$ , this problem is obviously decidable in polynomial time. But how can we decide whether two given finite structures are indistinguishable by expressions of  $\mathcal{F}$  without a bound on the degree? In this section, using the preceding results, we will show that this problem is also decidable in polynomial time.

For simplicity of presentation, we will work with the specific fragment  $\mathcal{C}(\bar{\pi})$  already used as an example fragment in Section 3.1. The generalization of the results in this section to the other fragments is left as an exercise to the reader.

Recall from Lemma 3 that expressions of  $\mathcal{C}(\bar{\pi})$  of degree at most  $k$  can only return pairs that belong to  $\text{paths}_k^{\mathcal{C}(\bar{\pi})}$ . Also recall from the discussion preceding that lemma that  $\text{paths}_k^{\mathcal{C}(\bar{\pi})}$  is the same as  $\text{paths}_k^{\mathcal{C}}$  and returns all pairs  $(a, b)$  of nodes such that  $b$  is reachable from  $a$  by a path of length at most  $2^k$ . In the following we will abbreviate  $\text{paths}_k^{\mathcal{C}}$  simply as  $\text{paths}_k$ .

Recall the notion of similarity up to depth  $k$  from Definition 7. We can define this notion equivalently through the following notion of simulation.

**Definition 27.** Let  $\mathcal{G}_1$  and  $\mathcal{G}_2$  be two structures with node sets  $V_1$  and  $V_2$  respectively. Let  $k$  be a natural number. Let  $\bar{Z} = (Z_0, Z_1, \dots, Z_k)$  and  $\bar{W} = (W_0, W_1, \dots, W_k)$  be tuples of relations with  $Z_i \subseteq V_1^2 \times V_2^2$  and  $W_i \subseteq V_2^2 \times V_1^2$  for  $i = 0, \dots, k$ . The pair  $(\bar{Z}, \bar{W})$  is called a *simulation up to depth  $k$*  from  $\bar{\mathcal{G}}_1$  to  $\bar{\mathcal{G}}_2$  if the following conditions are satisfied:

**Atoms** Assume  $(a_1, b_1, a_2, b_2) \in Z_i$ . If  $a_1 = b_1$ , then  $a_2 = b_2$ ; furthermore, for each  $R \in \Lambda$ , if  $(a_1, b_1) \in R^{\mathcal{G}_1}$ , then  $(a_2, b_2) \in R^{\mathcal{G}_2}$ .

**Composition Forth** Assume  $(a_1, b_1, a_2, b_2) \in Z_i$  with  $i > 0$ . Then for every  $c_1 \in V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_{i-1}(\mathcal{G}_1)$ , there exists  $c_2 \in V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_{i-1}(\mathcal{G}_2)$ , such that both  $(a_1, c_1, a_2, c_2) \in Z_{i-1}$  and  $(c_1, b_1, c_2, b_2) \in Z_{i-1}$ .

**Coprojection Forth** Assume  $(a_1, b_1, a_2, b_2) \in Z_i$  with  $i > 0$  and  $a_1 = b_1$  (whence  $a_2 = b_2$  by the Atoms condition). Then for every  $c_2 \in V_2$  with  $(a_2, c_2)$  in  $\text{paths}_{i-1}(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(a_1, c_1)$  in  $\text{paths}_{i-1}(\mathcal{G}_1)$  such that  $(a_2, c_2, a_1, c_1) \in W_{i-1}$ . Furthermore, for every  $c_2 \in V_2$  with  $(c_2, a_2)$  in  $\text{paths}_{i-1}(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(c_1, a_1)$  in  $\text{paths}_{i-1}(\mathcal{G}_1)$  such that  $(c_2, a_2, c_1, a_1) \in W_{i-1}$ .

**Reverse conditions** The above three conditions repeated, but with  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , and  $\bar{Z}$  and  $\bar{W}$ , exchanged:

**Reverse Atoms** Assume  $(a_2, b_2, a_1, b_1) \in W_i$ . If  $a_2 = b_2$ , then  $a_1 = b_1$ ; furthermore, for each  $R \in \Lambda$ , if  $(a_2, b_2) \in R^{\mathcal{G}_2}$ , then  $(a_1, b_1) \in R^{\mathcal{G}_1}$ .

**Reverse Composition Forth** Assume  $(a_2, b_2, a_1, b_1) \in W_i$  with  $i > 0$ . Then for every  $c_2 \in V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_{i-1}(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_{i-1}(\mathcal{G}_1)$ , such that both  $(a_2, c_2, a_1, c_1) \in W_{i-1}$  and  $(c_2, b_2, c_1, b_1) \in W_{i-1}$ .

**Reverse Coprojection Forth** Assume  $(a_2, b_2, a_1, b_1) \in W_i$  with  $i > 0$  and  $a_2 = b_2$  (whence  $a_1 = b_1$  by the Atoms condition). Then for every  $c_1 \in V_1$  with  $(a_1, c_1)$  in  $\text{paths}_{i-1}(\mathcal{G}_1)$ , there exists  $c_2 \in V_2$  with  $(a_2, c_2)$  in  $\text{paths}_{i-1}(\mathcal{G}_2)$  such that  $(a_1, c_1, a_2, c_2) \in Z_{i-1}$ . Furthermore, for every  $c_1 \in V_1$  with  $(c_1, a_1)$  in  $\text{paths}_{i-1}(\mathcal{G}_1)$ , there exists  $c_2 \in V_2$  with  $(c_2, a_2)$  in  $\text{paths}_{i-1}(\mathcal{G}_2)$  such that  $(c_1, a_1, c_2, a_2) \in Z_{i-1}$ .

The above definition of simulation corresponds to the notion of similarity we already have. We will prove this in the next Proposition 30. For that proof, we first need the following definition and two lemmas.

For two tuples of relations  $\bar{Z}$  and  $\bar{Z}'$  of the same length, we define their union  $\bar{Z}'' = \bar{Z} \cup \bar{Z}'$  in the obvious pointwise manner by  $Z_i'' := Z_i \cup Z_i'$  for each  $i$ . Similarly for two simulations  $(\bar{Z}, \bar{W})$  and  $(\bar{Z}', \bar{W}')$  up to the same depth, we can define their union as  $(\bar{Z} \cup \bar{Z}', \bar{W} \cup \bar{W}')$ . The proof of the following lemma is straightforward.

**Lemma 28.** *If  $(\bar{Z}, \bar{W})$  and  $(\bar{Z}', \bar{W}')$  are simulations from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k$ , then their union  $(\bar{Z} \cup \bar{Z}', \bar{W} \cup \bar{W}')$  is also a simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k$ .*

Since the three reversed conditions in the definition of simulation are completely symmetric to the first three conditions, we also have the following:

**Lemma 29.**  *$(\bar{Z}, \bar{W})$  is a simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  if and only if  $(\bar{W}, \bar{Z})$  is a simulation from  $\mathcal{G}_2$  to  $\mathcal{G}_1$ .*

We now state:

**Proposition 30.**  *$(\mathcal{G}_1, a_1, b_1) \preceq_k^{C(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)$  if and only if there exists a simulation  $(\bar{Z}, \bar{W})$  from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  such that  $(a_1, b_1, a_2, b_2) \in Z_k$ .*

*Proof.* The if-direction is immediately verified by induction on  $k$ . The only-if direction can also be proven by induction on  $k$ . The case  $k = 0$  is clear. Now assume  $(\mathcal{G}_1, a_1, b_1) \preceq_{k+1}^{C(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)$ . Then for each  $c_1 \in V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_k(\mathcal{G}_1)$ , there exists  $c_2 \in V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_k(\mathcal{G}_2)$ , such that both  $(\mathcal{G}_1, a_1, c_1) \preceq_k (\mathcal{G}_2, a_2, c_2)$  and  $(\mathcal{G}_1, c_1, b_1) \preceq_k (\mathcal{G}_2, c_2, b_2)$ . By induction, this is equivalent to the existence of  $c_2 \in V_2$  and of simulations up to depth  $k$   $(\bar{Z}_{\text{forth}}^{a_1, c_1}, \bar{W}_{\text{forth}}^{a_1, c_1})$  and  $(\bar{Z}_{\text{forth}}^{c_1, b_1}, \bar{W}_{\text{forth}}^{c_1, b_1})$  from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  such that  $(a_1, c_1, a_2, c_2) \in (Z_{\text{forth}}^{a_1, c_1})_k$  and  $(c_1, b_1, c_2, b_2) \in (Z_{\text{forth}}^{c_1, b_1})_k$ .

Furthermore, if  $a_1 = b_1$  (and  $a_2 = b_2$ ), then for every  $c_2 \in V_2$  with  $(a_2, c_2) \in \text{paths}_k(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(a_1, c_1) \in \text{paths}_k(\mathcal{G}_1)$  such that  $(\mathcal{G}_2, a_2, c_2) \preceq_k (\mathcal{G}_1, a_1, c_1)$ . By induction, and Lemma 29, this is equivalent to the existence of  $c_1 \in V_1$  and a simulation  $(\bar{Z}_{\text{coproj}_2}^{a_2, c_2}, \bar{W}_{\text{coproj}_2}^{a_2, c_2})$  up to depth  $k$  from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  such that  $(a_2, c_2, a_1, c_1) \in (W_{\text{coproj}_2}^{a_2, c_2})_k$ . Similarly, by the second part of the Coprojection Forth condition, for every  $c_2 \in V_2$  there exists  $c_1 \in V_1$  and a simulation  $(\bar{Z}_{\text{coproj}_1}^{c_2, a_2}, \bar{W}_{\text{coproj}_1}^{c_2, a_2})$  up to depth  $k$  from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  such that  $(c_2, a_2, c_1, a_1) \in (W_{\text{coproj}_1}^{c_2, a_2})_k$ .

All the above simulations can be extended up to depth  $k + 1$  by setting the  $k + 1$ st component to empty. The desired simulation is now obtained by taking the union of all these simulations (using Lemma 28), to which we add the single tuple  $(a_1, b_1, a_2, b_2)$  in the  $k + 1$ st component.  $\square$

For two tuples of relations  $\bar{Z}$  and  $\bar{Z}'$  of the same length, we say that  $\bar{Z} \subseteq \bar{Z}'$  if the inclusion holds pointwise, i.e.,  $Z_i \subseteq Z'_i$  for every  $i$ . Similarly, for two simulations  $(\bar{Z}, \bar{W})$  and  $(\bar{Z}', \bar{W}')$  up to the same depth, we define  $(\bar{Z}, \bar{W}) \subseteq (\bar{Z}', \bar{W}')$  if  $\bar{Z} \subseteq \bar{Z}'$  and  $\bar{W} \subseteq \bar{W}'$ .

We conclude:

**Proposition 31.** *Given two finite structures  $\mathcal{G}_1$  and  $\mathcal{G}_2$  and a natural number  $k$ , there is a unique simulation from  $\mathcal{G}_1$  and  $\mathcal{G}_2$  up to depth  $k$ , denoted by  $\text{Sim}(\mathcal{G}_1, \mathcal{G}_2, k)$ , that is maximal in that every other simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k$  is included in  $\text{Sim}(\mathcal{G}_1, \mathcal{G}_2, k)$ . Moreover,  $(\mathcal{G}_1, a_1, b_1) \preceq_k (\mathcal{G}_2, a_2, b_2)$  if and only  $(a_1, b_1, a_2, b_2) \in Z_k$  where  $\text{Sim}(\mathcal{G}_1, \mathcal{G}_2, k) = (\bar{Z}, \bar{W})$ .*

*Proof.* The maximal simulation equals the union of all simulations from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k$ . There are only finitely many such simulations since the structures are finite; their union is a simulation by Lemma 28. The proposition then follows from the previous proposition.  $\square$

For later use we note the following property of the maximal simulation, which follows immediately from the above proposition, Proposition 8, and Lemma 29:

**Lemma 32.** *The maximal simulation  $(\bar{Z}, \bar{W}) = \text{Sim}(\mathcal{G}_1, \mathcal{G}_2, k)$  up to depth  $k$  is monotonically decreasing, i.e., it satisfies  $Z_i \supseteq Z_{i+1}$  and  $W_i \supseteq W_{i+1}$  for  $0 \leq i < k$ .*

We next introduce an operator by which a simulation up to depth  $k$  can be refined up to depth  $k + 1$ .

**Definition 33.** Let  $\mathcal{G}_1$  and  $\mathcal{G}_2$  be two structures with node sets  $V_1$  and  $V_2$  respectively, and let  $k$  be a natural number. Let  $Z \subseteq V_1^2 \times V_2^2$  and  $W \subseteq V_2^2 \times V_1^2$ . We define  $\text{Refine}_{k+1}(Z, W)$  to be the pair  $(Z', W')$  where

- $Z'$  is the set of all tuples  $(a_1, b_1, a_2, b_2) \in Z$  satisfying the following two conditions:

**Composition Forth** For every  $c_1 \in V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_k(\mathcal{G}_1)$ , there exists  $c_2 \in V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_k(\mathcal{G}_2)$ , such that both  $(a_1, c_1, a_2, c_2) \in Z$  and  $(c_1, b_1, c_2, b_2) \in Z$ .

**Coprojection Forth** Assuming  $a_1 = b_1$ , then for every  $c_2 \in V_2$  with  $(a_2, c_2)$  in  $\text{paths}_k(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(a_1, c_1)$  in  $\text{paths}_k(\mathcal{G}_1)$  such that  $(a_2, c_2, a_1, c_1) \in W$ . Furthermore, for every  $c_2 \in V_2$  with  $(c_2, a_2)$  in  $\text{paths}_{i-1}(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(c_1, a_1)$  in  $\text{paths}_{i-1}(\mathcal{G}_1)$  such that  $(c_2, a_2, c_1, a_1) \in W$ .

- $W'$  is the set of all tuples  $(a_2, b_2, a_1, b_1) \in W$  satisfying the following two conditions:

**Reverse Composition Forth** For every  $c_2 \in V_2$  with  $(a_2, c_2)$  and  $(c_2, b_2)$  in  $\text{paths}_k(\mathcal{G}_2)$ , there exists  $c_1 \in V_1$  with  $(a_1, c_1)$  and  $(c_1, b_1)$  in  $\text{paths}_k(\mathcal{G}_1)$ , such that both  $(a_2, c_2, a_1, c_1) \in W$  and  $(c_2, b_2, c_1, b_1) \in W$ .

**Reverse Coprojection Forth** Assuming  $a_2 = b_2$ , for every  $c_1 \in V_1$  with  $(a_1, c_1)$  in  $\text{paths}_k(\mathcal{G}_1)$ , there exists  $c_2 \in V_2$  with  $(a_2, c_2)$  in  $\text{paths}_k(\mathcal{G}_2)$  such that  $(a_1, c_1, a_2, c_2) \in Z$ . Furthermore, for every  $c_1 \in V_1$  with  $(c_1, a_1)$  in  $\text{paths}_k(\mathcal{G}_1)$ , there exists  $c_2 \in V_2$  with  $(c_2, a_2)$  in  $\text{paths}_k(\mathcal{G}_2)$  such that  $(c_1, a_1, c_2, a_2) \in Z$ .

The two main properties of the Refine operator are:

**Lemma 34.** *Let  $\mathcal{S} = (\bar{Z}, \bar{W})$  be a simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k$ , let  $(Z_{k+1}, W_{k+1}) = \text{Refine}_{k+1}(Z_k, W_k)$ , and let  $\mathcal{T} = ((\bar{Z}, Z_{k+1}), (\bar{W}, W_{k+1}))$ . Then*

1.  $\mathcal{T}$  is a simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k+1$ .
2. Assume  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are finite. If  $\mathcal{S}$  is the maximal simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k$ , then  $\mathcal{T}$  is the maximal simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k+1$ .

*Proof.* The proof of the first property is immediate. To show maximality, consider the maximal simulation  $\mathcal{S}' = (\bar{Z}', \bar{W}')$  from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth  $k+1$ . We have to show that  $\mathcal{S}' \subseteq \mathcal{T}$ . Since  $\mathcal{S}$  is maximal up to depth  $k$ , we know that  $((Z'_0, \dots, Z'_k), (W'_0, \dots, W'_k))$ , which is a simulation up to depth  $k$ , is included in  $\mathcal{S}$ . So it remains to show  $(Z'_{k+1}, W'_{k+1}) \subseteq (Z_{k+1}, W_{k+1})$ . By Lemma 32, we have  $Z'_{k+1} \subseteq Z'_k \subseteq Z_k$  and  $W'_{k+1} \subseteq W'_k \subseteq W_k$ . Then by comparing the definition of simulation up to depth  $k+1$  with Definition 33, it is clear that  $(Z'_{k+1}, W'_{k+1}) \subseteq (Z_{k+1}, W_{k+1})$ , as desired.  $\square$

We can finally conclude:

**Theorem 35.** *The following problem is decidable in polynomial time: given two finite marked structures  $\bar{\mathcal{G}}_1$  and  $\bar{\mathcal{G}}_2$ , decide whether  $\bar{\mathcal{G}}_1 \equiv_{\mathcal{C}(\bar{\pi})} \bar{\mathcal{G}}_2$  holds.*

*Proof.* Let  $(Z_0, W_0)$  be the maximal simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to depth 0. So,  $Z_0$  consists of all tuples  $(a_1, b_1, a_2, b_2) \in V_1^2 \times V_2^2$  that satisfy the Atoms condition for  $i = 0$ , and  $W_0$  consists of all tuples  $(a_2, b_2, a_1, b_1) \in V_2^2 \times V_1^2$  that satisfy the Reversed Atoms condition for  $i = 0$ . Now for every natural number  $k > 0$ , define by induction  $(Z_k, W_k) := \text{Refine}_k(Z_{k-1}, W_{k-1})$ . By Lemma 34, in this way we obtain the maximal simulation from  $\mathcal{G}_1$  to  $\mathcal{G}_2$  up to ever increasing

depths. By Proposition 31 and Theorem 11, we have  $(a_1, b_1, a_2, b_2) \in Z_k$  iff  $(\mathcal{G}_1, a_1, b_1) \equiv_{\mathcal{C}(\bar{\pi})_k} (\mathcal{G}_2, a_2, b_2)$ .

For  $(\mathcal{G}_1, a_1, b_1) \equiv_{\mathcal{C}(\bar{\pi})} (\mathcal{G}_2, a_2, b_2)$  to hold, we need to verify whether  $(a_1, b_1, a_2, b_2) \in Z_k$  for *all* natural numbers  $k$ . Since  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are finite, and the sequence is monotonically decreasing (Lemma 32), there exists  $\ell$  such that  $Z_k = Z_\ell$  for all  $k \geq \ell$ . It thus suffices to compute  $(Z_k, W_k)$  for increasing  $k$  until no changes occur and check whether  $(a_1, b_1, a_2, b_2) \in Z_\ell$ . Denoting the maximum of the cardinalities of  $V_1$  and  $V_2$  by  $n$ , we have  $\ell \leq 2n^4$ , since in the worst case each iteration decreases one of  $Z_k$  or  $W_k$  by a single 4-tuple. Each iteration of the Refine operator is clearly computable in polynomial time. Thus the theorem is proved.  $\square$

## 7 Concluding remark

In our work, we have always included the identity relation and the three operations union, intersection and composition in the logics that we consider. It is an interesting topic for further research to see what happens if some of these operators are left out. In some aspects, the problems can change drastically. For example, consider the logic consisting only of composition and nothing else. Then indistinguishability of finite structures amounts to the equivalence problem for finite automata, which is PSPACE-complete [4], as opposed to the polynomial-time decidability we established for the fragments considered in this paper.

## References

- [1] RDF primer. W3C Recommendation, February 2004.
- [2] S. Abiteboul, P. Buneman, and D. Suciu. *Data on the Web: From Relations to Semistructured Data and XML*. Morgan Kaufmann, 1999.
- [3] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison-Wesley, 1995.
- [4] A.V. Aho, J.E. Hopcroft, and J.D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, 1974.
- [5] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, 2003.
- [6] M. Benedikt, W. Fan, and G. Kuper. Structural properties of XPath fragments. *Theoretical Computer Science*, 336(1):3–31, May 2005.
- [7] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.

- [8] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Springer, 1999.
- [9] G.H.L. Fletcher, D. Van Gucht, Y. Wu, M. Gyssens, S. Brenes, and J. Paredaens. A methodology for coupling fragments of XPath with structural indexes for XML documents. *Information Systems*, 34(7):657–670, 2009.
- [10] D. Florescu, A.Y. Levy, and A.O. Mendelzon. Database techniques for the World-Wide Web: A survey. *SIGMOD Record*, 27(3):59–74, 1998.
- [11] M.J. Franklin, A. Halevy, and D. Maier. From databases to dataspace: A new abstraction for information management. *SIGMOD Record*, 34(4):27–33, 2005.
- [12] S. Givant. The calculus of relations as a foundation for mathematics. *J. Autom. Reasoning*, 37(4):277–322, 2006.
- [13] V. Goranko and M. Otto. Model theory of modal logic. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, chapter 5. Elsevier, 2007.
- [14] M. Gyssens, J. Paredaens, D. Van Gucht, and G.H.L. Fletcher. Structural characterizations of the semantics of XPath as navigation tool on a document. In *Proceedings of the 25th ACM Symposium on Principles of Database Systems*, pages 318–327. ACM New York, 2006.
- [15] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, 2000.
- [16] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.
- [17] Robin Hirsch and Ian Hodkinson. *Relation Algebras by Games*. Elsevier, 2002.
- [18] L. Libkin. *Elements of Finite Model Theory*. Springer, 2004.
- [19] R.D. Maddux. The origin of relation algebras in the development and axiomatization of the calculus of relations. *Studia Logica*, 50(3/4):421–455, 1991.
- [20] R.D. Maddux. *Relation Algebras*. Elsevier, 2006.
- [21] N. Mamoulis. Efficient processing of joins on set-valued attributes. In *Proceedings ACM SIGMOD International Conference on Management of Data*, pages 157–168, 2003.
- [22] M. Marx. Conditional XPath. *ACM Transactions on Database Systems*, 30(4):929–959, 2005.
- [23] M. Marx and M. de Rijke. Semantic characterizations of navigational XPath. *SIGMOD Record*, 34(2):41–46, June 2005.



- [24] M. Marx and Y. Venema. *Multi-Dimensional Modal Logic*. Springer, 1997.
- [25] V.R. Pratt. Origins of the calculus of binary relations. In *Proceedings of the 7th IEEE Symposium on Logic in Computer Science*, pages 248–254, 1992.
- [26] A. Tarski. On the calculus of relations. *Journal of Symbolic Logic*, 6(3):73–89, 1941.
- [27] A. Tarski and S. Givant. *A Formalization of Set Theory Without Variables*, volume 41 of *Colloquium Publications*. American Mathematical Society, 1987.
- [28] J. van Benthem and J.A. Bergstra. Logic of transition systems. *Journal of Logic, Language and Information*, 3(4):247–283, 1994.
- [29] Y. Wu, D. Van Gucht, M. Gyssens, and J. Paredaens. A study of a positive fragment of path queries: expressiveness, normal form and minimization. *The Computer Journal*, 54(7):1091–1118, 2011.